Comparative genomic approaches to understanding *Achromobacter xylosoxidans*

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Pisut Pongchaikul

September 2015

UNIVERSITY OF LIVERPOOL

Acknowledgements

First of all, I am very grateful to immeasurable supports and encouragement from my supervisors – Dr Alistair Darby, Professor Craig Winstanley, and Dr Svetlana Antonyuk. Sincerely thanks all members of Darby's lab, in particular, Dr Ian Goodhead, for all of your kindness and lively environment in the office. Thanks all from Winstanley's lab. Thanks to Dr Pitak and Clinical Microbiology unit for clinical isolates and microbiological skill training. And thanks all worms died for my project.

Massive thanks all people in B231- all informaticians, Dr Richard Gregory, in particular. To, Jen (Dr Jennifer Kelly), thanks for your advise on pan-genome. To, Ian Wilson, many thanks for being nice to me, and for proofreading this thesis. To Sarah, thanks for being such a really nice friend.

Thanks all places where allow me to sit and write up this thesis. Thanks Chaba Chaba Thai restaurant for being my second home in Liverpool.

Special thanks Medical Scholar Programme for the opportunity to do intercalated degree. Thanks Mahidol-Chamlong Harinasuta PhD scholarship for an invaluable opportunity to come to Liverpool and to undertake such a great project. Thanks Dr Pansakorn, without your encouragement I would not have had a chance to do my PhD in Liverpool. Thanks PR601's members, especially Dr Ponpan, for my first experience in scientific research and for your mental support.

Last but not least, this pride is for my family, of course, my Mum and my Dad, who always boost me up. Thanks for your everlasting support. To a special person, Jitpisuth, you are everything to me – good friend, nice chef, travelling companion, French teacher, thesis proofreader, and wonderful girlfriend.

Declaration

I, myself, declare that all experiments described in this thesis have been carried out by me without any collaboration except for those indicated in acknowledgement and in the text. The works were produced in the Institute of Integrative Biology and the Institute of Infection and Global Health, the University of Liverpool.

Pisut Pongchaikul

Abstract

The investigation into the genome of the emerging pathogen reveals the genetic basis of the pathogen. In this study, *Achromobacter xylosoxidans* was used as an example for the genomic study of emerging pathogens. *A. xylosoxidans* is an emerging and opportunistic pathogen in patients with various underlying diseases, such as Cystic Fibrosis and cancers.

In chapter 3, the objectives of the study were to compare and evaluate the approach for *A. xylosoxidans* identification. A wide range of methods, including phenotypic test, 16S rDNA gene sequencing, MALDI-TOF, RAPD and MLST, were utilised to demonstrate the species identification. MALDI-TOF was considered as the most appropriate method due to the least time consuming. The application of multiple approaches to identify *A. xylosoxidans* was suggested.

In chapter 4, comprehensive genomic feature of *A. xylosoxidans* has not been elucidated. The objective of this study was to use comparative genomic tool to investigate genomic feature of *A. xylosoxidans*. The analysis revealed the opened pan-genome of the species. The core genome accounted for approximately 50% of the size of the genome. Furthermore, the analysis revealed recombination events in the core genome of the species. Interestingly, phylogenetic relationships demonstrated global distribution of the species without geographical structure. This study provided pan-genome structure of the species, allowing for studies of genetic exchange mechanism in the species.

In chapter 5, the main objective of the chapter was to investigate the antibiotic resistance and genes associated with the resistance in the species. The whole genome sequence and bioinformatics were used to search for genes associated with antibiotic resistance phenotypes. Single-molecule real time (SMRT) sequencing was also used to investigate the integron. The analysis revealed conserved RND-type efflux transporters across the species. The complete genome sequence revealed class 1 integron carrying IMP-14 on the chromosome of multidrug resistant isolates. This study demonstrated the identification of antibiotic resistance genes using bioinformatics and SMRT sequencing.

In chapter 6, the objective was to investigate the virulence of *A. xylosoxidans* in organism model. The great moth larva *Galleria mellonella* was used to test for the virulence of *A. xylosoxidans*. Bioinformatics analyses were conducted to predict genes associated with virulence determinants. The analysis resulted in hydrolase-containing protein and PscD type III secretion system predicted as virulence determinants based on *G. mellonella* infection model, paving the way for further studies.

This study demonstrated the application of whole genome sequencing and bioinformatics to understand the biology of the emerging pathogen. The identification of the genome of the emerging pathogen is required for the facilitation of prevention and therapeutic in the future.

Glossary

°C	Degree celcius
DNA	Deoxyribonucleic acid
FEV_1	Forced expiratory volume in one second
g	Gram
h	Hour
L	Litre
$MgCl_2$	Magnesium Chloride
m	Metre
µ	Micro
mM	Millimolarity
min	Minute
M	Molarity
mol	Mole
n	Nano
OD	Optical density
pmole	Picomole
RNA	Ribonucleic acid
rDNA	Ribosomal DNA
U.K.	United Kingdom
U.S.	United States of America

Table of Contents

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 7

List of Figures

List of Tables

Chapter 1

General Introduction

Infectious diseases are still having a remarkable impact on the quality of life around the world. A recent survey by World Health Organization (WHO) shows that infectious diseases are in the top ten causes of death (Figure 1.1). The largest causes of death are ischaemic cardiac diseases, stroke and chronic obstructive pulmonary disease (COPD). However, the 4th, 6th and 7th rank are infectious diseases: lower respiratory tract infection, Acquired Immune Deficiency Syndrome (AIDS) and diarrhoea, respectively. Focusing on countries grouped by national income, infectious diseases are in the three highest causes of death in low-income countries (Figure 1.2).

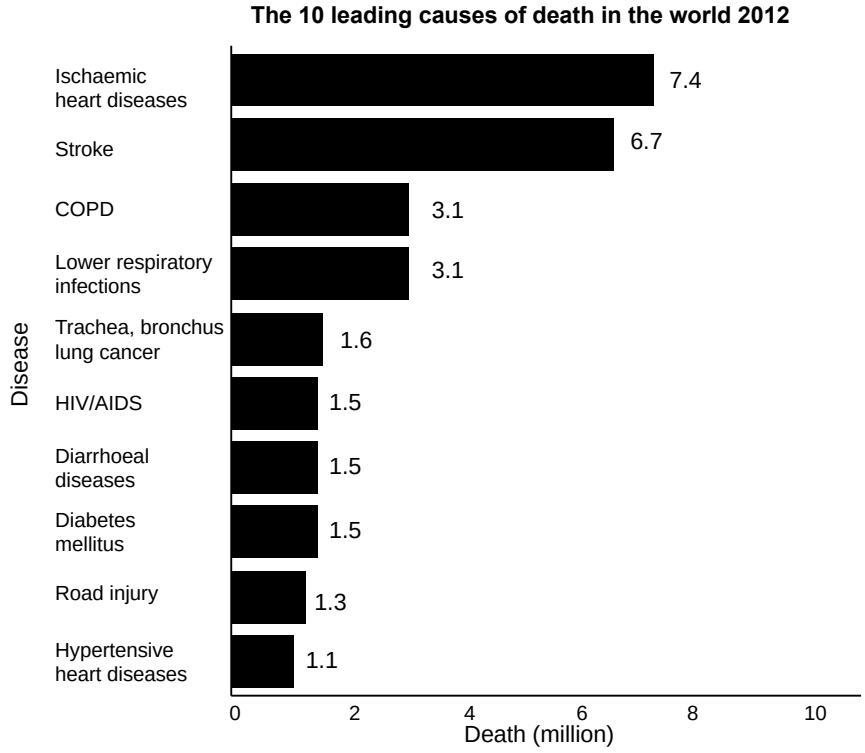


Figure 1.1: The ten highest causes of death according to a world's population survey by WHO in 2012 (http://www.who.int/mediacentre/factsheets/fs310/en/)

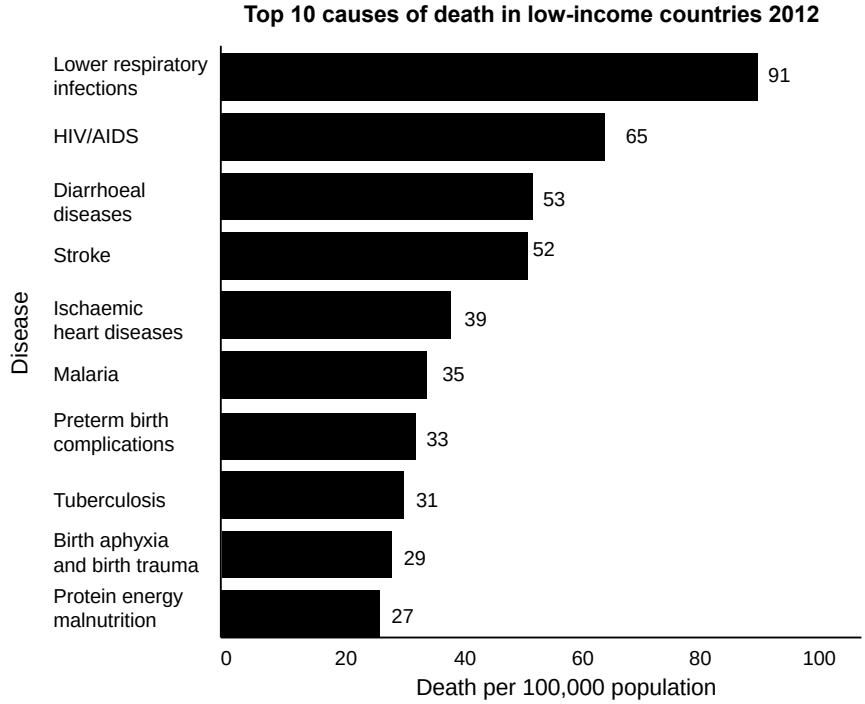


Figure 1.2: The ten highest causes of death in low-income countries according to a surveillance by WHO in 2012 (http://www.who.int/mediacentre/factsheets/fs310/en/index1.html)

Life-threatening infectious diseases such as poliomyelitis and small pox have been eliminated using vaccination. However, many of infectious diseases still remain with high virulence, such as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). Bacterial infections still remain, and importantly, they are caused by antibiotic resistant pathogens. In addition, infections by emerging pathogens have become a troublesome factor for clinical medicine.

The introductory part of this thesis will present the general situations of emerging infectious diseases and hospital-acquired infections. Furthermore, the impact of the diseases and prevention strategies will be addressed to state the importance of conducting the comprehensive research on emerging infectious diseases.

1.1. Emerging infectious diseases and hospital-acquired infections

1.1.1. Emerging pathogens

As infectious diseases remain one of the major causes of death in the global survey (Figure 1.1 and 1.2), emerging infectious diseases are becoming a challenge due to the lack of information about the pathogens, for example, the routes of transmission and the profile of drug resistance. Three possible mechanisms of how emerging infectious diseases occur have been proposed by Engering, Hogerwerf & Slingenbergh (2013):

1. New hosts: The pathogen moves from one species to another species by accident. The driving factor of this event is usually a close contact between hosts such as human and animals. Common examples of the driving factors are the consumption of raw wild meat and the contact between human and domestic animals. The jump from one species to another species requires the adaptation of the pathogen in order to successfully survive in a new host. For example, the transmission of *Streptococcus suis* from pigs to human leads to serious infections, including, septicaemia and meningitis, in human (Wertheim *et al.*, 2009). This demonstrates a particular type of the transmission of pathogens from animals to humans, called 'zoonosis'. Another example of a bacterial zoonotic infection is cat scratch disease caused by *Bartonella* species (Chomel & Kasten, 2010).

2. New phenotypes: This event happens when bacteria that have already existed in human. The external disturbances make bacteria adapt to survive in the body of the hosts, by expressing new phenotypes that enhance the viability under the changed environment. The common disturbances are the use of antibiotic and vaccination to eliminate the pathogens. Antibiotic resistance is, generally, expressed by emerging pathogens, as a new trait, to survive in the harsh environment of antibiotic use. The outbreak of drug resistance *Acinetobacter*

baumannii in nursing homes (Sengstock *et al.*, 2010) and the outbreak of vaccine resistance *Streptococcus pneumoniae* in the US (Brueggemann *et al.*, 2007) are the example of the expression of new trait of emerging pathogens.

3. New geography: This mechanism is considered when the pathogens, which have already been endemic in a particular area, spread to different area on account of the change in physical factors such as temperature and land use. The spread of dengue haemorrhagic fever in Southeast Asian countries is one example (http://www.who.int/mediacentre/factsheets/fs117/en/). Climate change is usually a driving factor that triggers an outbreak of emerging pathogen. Global warming has a major effect on the emergence of diseases, particularly vector-borne diseases. For instance, the outbreak of dengue haemorrhagic fever in Europe was driven by climate change (Bouzid *et al.*, 2014). In addition, the expansion of land use also contributes to the appearance of emerging infectious diseases. For instance, the extension of land use in th US between 1970s and 1980s accounted for the distribution of Lyme disease, which is caused by *Borrelia burgdorferi*, from northeastern area to the other area (Ciesielski *et al.*, 1988).

1.1.2. Hospital-acquired infection problem

The emergence of rare pathogens and multidrug-resistant pathogens increasingly poses problems for clinical service in the hospital worldwide. Hospital-acquired infections, also known as ‘nosocomial infections’, cause a major problem to healthcare providers. The term ‘hospital-acquired’ infections usually refers to infections that develop, at least, 48 hours after hospital admission (Falcone *et al.*, 2011). Hospitals are places where emerging pathogens are often detected. For example, the emergence of *Enterococcus faecium* in the hospital is driven by the antibiotic resistance phenotype (Weber & Gold, 2003). *Elizabethkingia meningoseptica* has also been emerged as hospital-associated pathogen on

account of its antibiotic resistance and its adaptability to a new environment (Jean *et al.*, 2014).

According to a survey from Center for Disease Control and Prevention of the US, severe hospital-associated infections occur in 1 in every 25 hospitals (http://www.cdc.gov/HAI/surveillance/index.html). In European countries, the European Centre for Disease Prevention and Control survey reports that approximately 7% of the patients had hospital-associated infections and 35% of patients, who have hospital-acquired infections, are given at least one antibiotic.

The most common types of nosocomial infections are the respiratory infection, surgical site infection, gastrointestinal tract infection, and urinary tract infection (Zarb *et al.*, 2012). Medical devices can be the reservoirs for pathogens and help them spread to other patients. Considering the causes of infection, medical device-associated infection contributed approximately 25% of nosocomial infections in the US.

1.1.2.1. Hospital-acquired respiratory infections

Hospital-associated respiratory tract infections, in particular pneumonia, cause a number of problems to hospitalised patients. Most of the hospital-acquired pneumonia are associated with medical devices, especially in critical care unit. Endotracheal intubation is the most common cause of pneumonia in hospitals. The intubation provides the opportunity of lung infection via four possible routes (Kalanuria, Zai & Mirski, 2014): (1) micro-aspiration during the intubation, (2) biofilm formation of bacteria inside the tube during the intubation, (3) accumulation of secretion around the tube where bacteria can inhabit, and (4) minor trauma of muco-ciliary clearance system during the intubation. The three leading causes of hospital-acquired respiratory tract infections are *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Escherichia coli*, and, importantly, most of them are antibiotic-resistance strains (Jones, 2010). Some emerging pathogens, such as *A. xylosoxidans*, are also causes of hospital-acquired infection, especially in Cystic Fibrosis infection (Kanellopoulou *et al.*, 2004; Rønne Hansen *et al.*,

2006; De Baets *et al.*, 2007; Hansen *et al.*, 2010; Lambiase *et al.*, 2011; De Baets *et al.*, 2013). Preventing the transmission of pathogens to patients is essential because most microbes that circulate in the hospital are antibiotic-resistance strains. Importantly, causing pathogens of hospital-acquired respiratory tract infections can be transmitted from person-to-person via air-borne route. The guideline from British Society for Antimicrobial and Chemotherapy recommends hand hygiene and the sterilisation of ventilators as a preventive guideline for hospital-acquired respiratory tract infections (Masterton *et al.*, 2008).

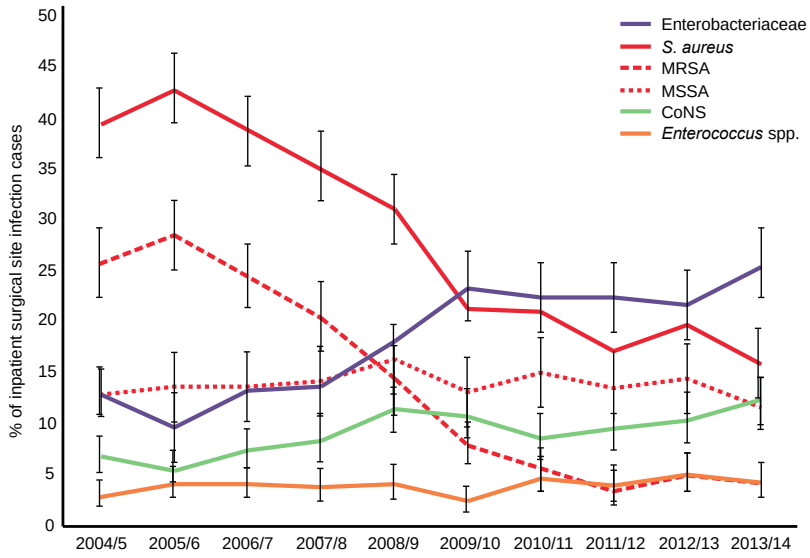


Figure 1.3: Important microorganisms in surgical site infection, reported by NHS hospital in the U.K. A graph is taken from 'Surveillance of Surgical Site Infection in NHS Hospitals in England' (Public Health England, 2014)**.**

1.1.2.2. Hospital-acquired surgical site infections

Infection near a surgical site is also concerning, because this type of infection accounts for approximately 15% of overall hospital-acquired infections (Watanabe *et al.*, 2008). By definition, the surgical site infection is the infection that occurs near or at surgical incision region within 30 days after the operation. The incision of the skin allows bacteria to get into the body. In many cases, bacteria that colonise near the incision area serve as a cause of infections. For instance, *S. aureus*, in particular Methicillin-resistance strains, are commonly found at the surgical cut site (Manian *et al.*, 2003). According to a report by

Public Health England, the prevalence of infections by Methicillin-resistance *S. aureus* gradually decreases from 40% to 16% of cases of surgical site infections, whereas the prevalence of infections caused by *Enterobacteiaceae* increases from 10% to 26% of cases of surgical site infections (Figure 1.3). Preventive strategies for surgical site infections include disinfecting skin's microbe before surgical operation and preventing staff-to-patient transmission during post-operative period. However, certain underlying conditions, such as diabetes mellitus, obesity and smoking, increase the risk from being infected with those pathogens (Reichman & Greenberg, 2009).

1.1.2.3. Hospital-acquired gastrointestinal tract infections

Gastrointestinal infection during hospital admission is one of the most common hospital-acquired infections (http://www.cdc.gov/HAI/surveillance/index.html). *Clostridium difficile* is recognised as a key organism that contributes to a remarkable problem in health care system (Dubberke & Olsen, 2012). The infection is associated with the lack of gut commensals due to an overuse of antibiotic (Bignardi, 1998). *C. difficile* infection can be transmitted via air-borne route. Best *et al.* (2010) revealed the presence of *C. difficile* in aerosol around symptomatic *C. difficile* patients, allowing the airborne dissemination of the pathogen. Although nosocomial gastrointestinal infection is life threatening (Parashar *et al.*, 2006; Marra *et al.*, 2007), the infection and transmission of the pathogens is preventable. The principal practice is to decrease the risk of transmission from person to person. The prevention using non-medication procedures, including education, hygiene and isolation, reduces the incidence of hospital-associated *C. difficile* infection (You *et al.*, 2014). Furthermore, decreasing the overuse of antibiotic in hospitalised patients can also reduce the number of *C. difficile* infection cases (Valiquette *et al.*, 2007). Recently, the administration of the probiotics decreases the prevalence of *C. difficile* infection (McFarland, 2006). Faecal transplantation has also been proposed as an new treatment for *C. difficile* infection (Gens, Elshaboury & Holt, 2013). Nosocomial viral infections are more dominant than bacterial infections in children. Rotavirus and norovirus are considered as major aetiologic factors of nosocomial

gastrointestinal infection in children (Oh, Gaedicke & Schreier, 2003). These viruses can be transmitted from person to person via fecal-oral route. It was also reported that viral particles were found in stool samples of patients who had gastroenteritis symptoms (Brady *et al.*, 1989). However, the infectivity of virus can be reduced by applying disinfectants on the surface whence the viruses are present (Sattar *et al.*, 1983, 1994).

1.1.2.4. Hospital-acquired urinary tract infections

Urinary tract infection is a common problem in post-operation patients due to the retaining of urinary catheter. Urinary catheter is most frequently used medical device, accounted for 17% in Europe (Zarb *et al.*, 2012) and 23% in the U.S. (Magill *et al.*, 2014). Generally, bacteria colonise around the distal urethra area because the upward migration of the bacteria is prevented by the outward flow of urine. With the insertion of catheter, flow of urine through the urethra is restricted, allowing for the upward migration of the bacteria. The presence of bacteria in the urinary bladder can follow these three possible ways: (1) the tip of the catheter pushes bacteria from the distal urethra to the bladder, (2) the bacteria gradually migrate through the catheter surface into the bladder, and (3) the drainage bag is contaminated by the bacteria during preparing process (Barford & Coates, 2009). The most common cause of catheter-associated infection is *E. coli* (Wagenlehner, Weidner & Naber, 2005). Other species such as *Enterococci spp.*, *P. aeruginosa*, and *A. baumannii* (Wagenlehner, Weidner & Naber, 2005; Djordjevic *et al.*, 2013) are also considered as causes of catheter-associated infections. An ideal solution of this problem is to avoid using catheter or to decrease the duration of catheter intention (Hooton *et al.*, 2010). Unfortunately, 74% of hospitals in the U.S. have not monitored the duration of catheterisation (Saint *et al.*, 2008). Alternative ways to reduce the infection are to use antibiotic-coated catheter, to monitor the bladder using ultrasonography, and to prescribe prophylactic antibiotic after catheter removal (Hooton *et al.*, 2010).

1.1.3. The burden of hospital-acquired infection and prevention strategies

Hospital-acquired infections also cause a socio-economic burden. In the U.S., approximately two million patients suffer from hospital-acquired infections each year (Reed & Kemmerly, 2009). Hospital-acquired infections increase the length of hospital stay, leading to increase the probability of receiving pathogens circulating in the hospital. Nathwani *et al.* (2014) revealed that extra-period in the hospital increased the chance of infection from multidrug-resistance *P. aeruginosa*, compared to non-resistant strains. Considering economic burden, a meta-analysis conducted in the U.S. reports that hospital-acquired infections cost 603 to 65,245 USD per case, and cost 8.3 – 11.5 billion USD per annum (Zimlichman *et al.*, 2013). In the U.K., the infection costs approximately one billion GBP per annum (Plowman, 2000). Infection at surgical site contributes to the largest partition (33.7%), and catheter-associated infection contributes to the smallest partition (<1%). However, several studies showed that both cost and cases of hospital-acquired infection decreased after infection control and prevention have been launched (Reed & Kemmerly, 2009; Zimlichman *et al.*, 2013).

As a result, some preventive strategies have been established. There are staffs responsible for infection control, have been set up in hospitals in many countries such as the U.K. (Plowman, 2000) and Japan (Morikane, 2012). Briefly, the roles of infection control team are (1) to educate hospitals' staffs and public via campaigns such as hand cleaning, (2) to support hospital's staff in infection control's practice, and (3) to survey infection information in the hospital. To educate healthcare staff is the most important strategy. Danzman *et al.* (2013) reported that healthcare providers are important carriers of nosocomial pathogens. Direct contact by healthcare staffs allows for the transmission of the pathogens from health care providers to patients. Hand hygiene is, therefore, an appropriate and effective practice for health care workers. General cleaning of the surfaces, that are contacted by hospital's staffs and patients, in the hospital can also prevents the transmission of the pathogens (Dancer, 2014). Furthermore, guideline for each type of nosocomial infection such as Surgical Care Improvement Projects (SCIP) for surgical site infection (Bratzler & Hunt, 2006)

has been launched in order to reduce the prevalence of hospital-acquired infections. Apart from pathogen transmission, another problem troubling hospital-acquired infections is the resistance of pathogens to antimicrobial agents.

1.2. Antibiotic resistance

1.2.1. Antibiotic resistance: a multi-dimensional problem

Antimicrobial resistance is a major problem worldwide (http://www.who.int/drugresistance/documents/surveillancereport/en/). Many common pathogens have become resistant to effective antibiotics, for example, carbapenem-resistant *Klebsiella pneumoniae*, fluoroquinolone resistant *E. coli*, 3rd generation cephalosporin resistant *Neisseria gonorrhoeae*, and methicillin resistant *S. aureus*. Innate resistance to antibiotic has been found in some pathogens, for example, *A. xylosoxidans* (Hu *et al.*, 2015). In fact, this problem is, now, approaching the situation where only 'last resort' drugs are available.

1.2.1.1. Causes of antibiotic resistance

Antibiotic resistance causes a huge problem to infections, especially hospital-acquired infections. Antibiotics with better efficacy are needed for resistant bacteria treatment. Since the discovery of penicillin (Fleming, 1929), the researches for antibiotics have been on-going. The use of antibiotic to treat bacterial infection has been considered a major cause of drug resistance in bacteria. The first report of clinically important antibiotic resistance dated back to 1942 when penicillin-resistant *S. aureus* was described (Rammelkamp & Maxon, 1942). To develop to be antibiotic-resistant species, bacteria depict a natural selection theory that organisms with a positive selection under a particular environment can survive and reproduce. In other words, the bacteria that can maintain their fitness can survive.

The emergence of antibiotic resistant strains follows natural selection driven by the use of antibiotic as an evolutionary pressure (Blázquez, Oliver & Gómez-Gómez, 2002). The development of antibiotic resistance can be considered as an intrinsic development and an extrinsic development. For intrinsic process, the mutation of drug target-associated genes usually plays a role in this process. Kolar *et al.* (2001) revealed that cephalosporin-resistant bacteria increased as the use of cephalosporin increased, while resistant bacteria decreased when cephalosporin use is restricted. Likewise, a 13-year study in Taiwan showed the relationship between drug-resistance development and antibiotic use in the hospital (Hsueh, Chen & Luh, 2005). For extrinsic factor, the resistance can be transferred between the bacterial species. The observation of transferrable antibiotic resistance has been reported since 1959 (Ochiai *et al.*, 1959) that horizontal gene transfer plays the role in the resistance of bacteria to the antibiotics (Maiden, 1998). The horizontal gene transfer, also known as lateral gene transfer, is a transmission of genes that send genetic element to other bacteria via the movement of mobile genetic elements (Frost *et al.*, 2005). This type of the transfer of resistance gene has a huge impact on clinical microbiology because it demonstrates that resistance genes can be transferred not only to their offspring, but also to other bacteria. Taking all into consideration, the major cause of the distribution of antibiotic resistant pathogens is the inappropriate use of antibiotic. This also has a consequent impact on health financing.

1.2.1.2. Economic burden of antibiotic resistance

Antibiotic resistance does not only cause clinical burden, but it also causes the consequent economic burden. In most cases, newer and more effective antibiotics are more expensive than first-line antibiotics. In pandemic area of drug resistant pathogens, doctors replace first-line antibiotics with higher potent antibiotics. For instance, in patients with otitis media (the inflammation of the middle ear), amoxicillin-clavulanic acid is prescribed instead of amoxicillin alone. This costs six times higher than the first-line antibiotic prescription. It also results in the establishment of the resistance to more potent antibiotics in community (Howard, 2004). Although there has been an investment in the

strategies to solve an antibiotic resistance problem, such as increasing research funding for new antibiotics discovery, the problem still remains because of the existing inappropriate use of the antibiotic. Preventive strategy can be an appropriate solution for this problem. The prevention for antibiotic resistance does not only reduce the severity of the problem, but also reduce the cost spent on newer antibiotics. The investment in the preventive strategies for antibiotic resistance is less than the investment in the prescription of higher antibiotics and new antibiotic research. For instance, using of gowns and gloves of healthcare providers reduced a number of vancomycin-resistant *Enterococci*, compared to using gloves alone (Puzniak *et al.*, 2004). In addition to the cost of drug, infection from resistant microbes prolongs the length of hospital stay. Mauldin *et al.* (2010) addressed a significantly higher amount of additional hospitalisation cost spent on drug-resistant infection (29.3 %), compared to drug-susceptible infection. Taken together, it is important to note that antibiotic resistance problem affects the whole public health system, including biomedical research, healthcare service, and public health administration.

1.2.2. Antibiotic resistance in hospital-acquired emerging infection

Hospitals are important places where antibiotic resistant bacteria can be developed as a result of a high number of antibiotics used (Gold & Moellering, 1996). Multidrug-resistant pathogens can circulate in hospital via hospital staffs and infected patients (Danzmann *et al.*, 2013). This, consequently, leads to the spread of drug-resistant pathogens in the hospital, as exemplified by the local circulation of multidrug resistance *P. aeruginosa*, driven by hospital' staffs, in provincial hospital in Thailand (Kiddee *et al.*, 2013).

Clinically-naïve bacteria can often become human pathogens, often by expressing drug resistance, while they are circulating in the hospital. Some bacteria can transform from drug-susceptible strains to drug-resistance strains or some are accessing to the hospitals with intrinsic drug resistance. The emergence of methicillin-resistant *S. aureus* exemplifies the conversion into emerging pathogen by the expression of antibiotic resistance trait. *Enterococci* species

have adapted from commensals in human gut to be an emerging pathogen in nosocomial infection due to the intrinsic resistance to commonly used antibiotics (Morrison, Woodford & Cookson, 1997). *Enterococci* are recently resistant to linezolid, which is one of the top-line antibiotics (Auckland *et al.*, 2002). *Stenotrophomonas maltophila* changes its susceptibility from non-resistance to resistance due to antibiotic pressure in the hospital (Looney, Narita & Mühlemann, 2009). Vartivarian *et al.* (1994) demonstrated that hospital isolates of *S. maltophila* became more resistant to major antibiotics, including fluoroquinolones, co-trimoxazole and cephalosporins over 11 years of the study, between 1981 and 1992. *A. xylosoxidans* have already been considered emerging pathogen due to the intrinsic resistance (Hu *et al.*, 2015). Another example is that *Elizabethkingia meningoseptica* has become an emerging pathogen due to the multidrug resistance to bata-lactam antibiotics, particularly, carbapenems (González & Vila, 2012; Jean *et al.*, 2014). Virulence of several emerging pathogens has not been elucidated. However, it is noteworthy that most of emerging hospital-acquired pathogens cause serious infections due to their drug-resistance phenotypes.

Multidrug resistance of the emerging opportunistic pathogens also exemplifies the problem of nosocomial infection. Patients, who have a high risk for infection, such as cancer, diabetes and immune-compromised condition, are prone to be infected by the microbes. An example of emerging nosocomial infection is *Asaia lannensis* infection in children with heart disease during hospitalisation (Juretschko, Beavers-May & Stovall, 2010). This species showed pan-antibiotic resistance to most antibiotics for Gram-negative bacteria such as beta-lactams and fluoroquinolones. Respiratory tract infection caused by *A. xylosoxidans* in epiglottic cancer also illustrates the role of emerging pathogens in opportunistic infection (Priyamvada, 2014). Resistance to antibiotic in some pathogens can be predicted because those resistance genes are carried as intrinsic genes. However, acquired resistance is an adaptation process driven by antibiotic use (Blázquez, Oliver & Gómez-Gómez, 2002). Therefore, the regulation and the use of antibiotic in certain region can indicate the prevalence of antibiotic resistant infection in the region.

1.3. Situation of antibiotic resistance in countries where are associated with this study: The U.K. and Thailand

1.3.1. Antibiotic resistance problem in the U.K.

The United Kingdom has faced serious antibiotic resistance problem since the existence of methicillin-resistant *S. aureus* (Barber, 1961). Major causes of antibiotic resistance problem in the U.K. are from both internal factor and external factor. For internal cause, improper prescription is a key that accelerates antibiotic resistance. Approximately 80-90 % of antibiotic prescriptions happen in primary and secondary health care unit (Shallcross & Davies, 2015). An observation between 1995 and 2011 reported that some antibiotics were prescribed in primary health care centres. For instance, 51% of patients with self-limited cough and cold were prescribed antibiotics. Situation in secondary health care unit is worse than the one in primary units. The surveillance reported a significant increase in the use of broad-spectrum antibiotics, such as co-amoxiclav and carbapenem (Hawker *et al.*, 2014). To deal with the problems that need intensive follow-up, antimicrobial stewardship surveillance has commenced to improve the quality of prescription and to encourage reasonable prescription (Ashiru-Oredope & Hopkins, 2013).

For external cause, antibiotic resistance in England is shown to be influenced by pathogens from the outside. Studies showed that antibiotic resistance of enteric fever in East London was associated with overseas travelling (Reddy *et al.*, 2011; Dave *et al.*, 2015). Studies also demonstrated that the majority of patients (92%) infected with antibiotic resistance *Salmonella*, had just been back from travelling abroad. The same study revealed that the antibiotic resistance of bacteria remained over seven years of the survey, between 2005 and 2012 (Dave *et al.*, 2015).

1.3.2. Antibiotic resistance problem in Thailand

It has been acknowledged that antibiotic resistance affects public health in Asian countries (Kim *et al.*, 2013; Chewapreecha *et al.*, 2014). Thailand is one of the most famous destinations for travellers. Thailand has been reported as one of Asian countries that have confronted the antibiotic resisance challenge. A recent study in 2010 revealed that the resistance challenge in Thailand caused about 39,000 deaths and costed up to 112 million pounds sterling (GBP) (exchange rate: 54 Thai baht per GBP) (Sumpradit *et al.*, 2012). The access to antibiotic in Thailand has not been effectively controlled according to public health situation. For example, people in rural area prefer to purchase medicine from a chemist rather than to visit a medical doctor due to the imbalance of doctor-to-patient ratio and the inaccessibility of hospital visit. People can purchase antibiotic from the chemist with no need of medical prescription. This problem also happens in big cities, such as Bangkok. Several studies were shown that chemists distributed several antibiotics (Thamlikitkul, 1988; Apisarnthanarak *et al.*, 2008). A study in Patumthani province suggested the positive correlation between the rise of antibiotic misuse and the resistance of antibiotic use (Apisarnthanarak & Mundy, 2008). Inappropriate distribution of antibiotic through the chemist is considered as a major factor that triggers the antibiotic resistance (Apisarnthanarak *et al.*, 2008; Apisarnthanarak & Mundy, 2008). Furthermore, tertiary health care units and teaching hospitals experience antibiotic resistance's problems. Antibiotics are always given to patients who are hospitalised in order to prevent post-operative infection (Apisarnthanarak *et al.*, 2006). Surprisingly, the abuse of antibiotics is seen in the unit of internal medicine, which is specialised in the use of antibiotics (Kusuma Na Ayuthya *et al.*, 2003; Apisarnthanarak *et al.*, 2006). Although it has been recognised that the cause of antibiotic resistance is the overuse and inappropriate use of antibiotic, samples' collection and antibiogram cannot reveal the transmission of antibiotic resistance genes and the distribution of drug resistant pathogens. Whole genome sequencing is, therefore, an approach to reveal the genetic basis of resistance mechanisms and the molecular epidemiology of resistant pathogens.

1.4. Whole genome sequencing in clinical microbiology

The study of epidemiology of infectious diseases historically depends on the number of symptomatic cases in particular location and time. This information has been used to build the inference to the transmission of pathogens. However, the classical data collection is not always enough for the investigation for the identity and the virulence of the pathogens. The application of whole genome sequencing on microbiological studies paves the way for the understanding of molecular epidemiology of pathogens, antibiotic resistance-associated genes, and virulence-associated genes (Grad & Lipsitch, 2014).

1.4.1. Next-generation whole genome sequencing

DNA sequencing is a groundbreaking invention of the 20th century, allowing for biological studies in molecular level. The pioneering sequencing was invented by Sanger and his colleagues, using chain-terminating dideoxyribonucleotides (Sanger, Nicklen & Coulson, 1977). The current generation of DNA sequencers was introduced with an effort to improve the efficiency of DNA sequencing by reducing time and costs (Margulies *et al.*, 2005). There are two next-generation sequencing platforms applied in this study: Illumina® platform and Pacific Biosciences® platform.

Table 1.1: Next-generation sequencing platform used in this study (Loman *et al.*, 2012; Radford *et al.*, 2012)

Sequencing technology	Platform	Run time	Read length (bases)	Gb per run	Strength	Weakness
Sequencing by synthesis	Illumina HiSeq	11 days	2 × 100	600	• High yield per run • Cost effective	• Long run time • Short read lengths
SMRT technology	Pacific Biosciences RSII	4 hours	10^3-10^4	3 per day	• Long read lengths • Short run time	• High error rate

Illumina's sequencing is performed by fragmenting genomic DNA and ligating the fragmented sequences with adapters (Figure 1.4) (Mardis, 2013). Following the preparation of the DNA library, the adapter-ligated DNAs are loaded into a flow cell to attache to pre-coated primers, which are complimentary to the

adapters, on the surface. This platform uses a unique process called ‘bridge amplification’ (Figure 1.4). The free ends of the DNA fragment bend over to bind to a nearby complementary adapter, forming clusters of DNA on the slide. The sequencing process starts by the addition of fluorescently labelled dNTP. The addition’s rate is one base at a time because of the presence of a blocking group at 3’-OH preventing polymerisation reaction. The image sensor detects the wavelength and the intensity of fluorescent signals from the synthesised sequences, allowing for the generation of the nucleotide sequence. After the cleavage of fluorescent dye and deblocking 3’-OH end, the synthesis of nucleotides can carry out. The reaction of nucleotide synthesis happens for up to 300 bp before the synthesis of the second sequence commences. This process is called ‘sequencing by synthesis’ (Mardis, 2013). Illumina sequencing has been claimed to be an accuracy and appropriate approach for high throughput; however, the sequencing has also been reported for the nucleotide substitution mistake over GC rich region (Dohm *et al.*, 2008).

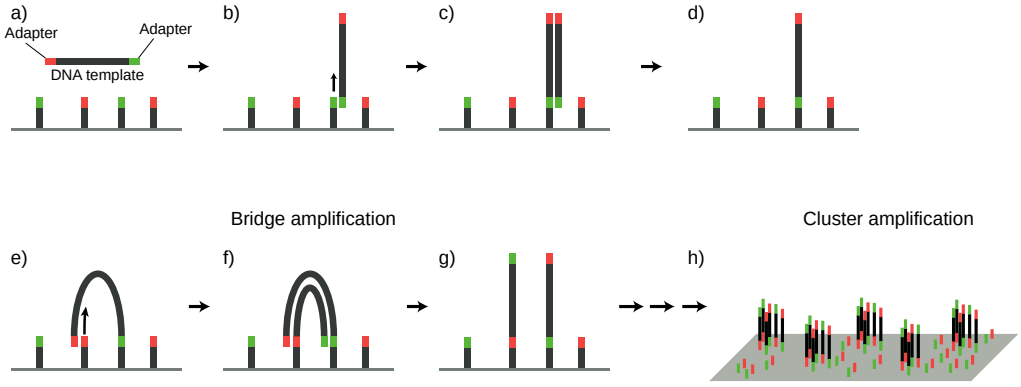


Figure 1.4: DNA sequencing by Illumina® technology. During annealing, adapter-tagged single-stranded DNAs bind to a complementary primer (b). The replication then occurs (c-d). A free end of amplified sequences bind to complementary primer close to them during annealing (e), allowing for “cluster amplification” using bridge amplification (f-h).

Single-molecule, real-time sequencing technology (SMRT) is on the basis of the undisturbed synthesis of single nucleotide sequence in each sequencing unit (Eid

et al., 2009). The double-stranded DNA template of the sequencing is bound with adapters to achieve circularly hairpin formation. Then, the prepared DNA strands were trapped by the polymerases. Unlike Illumina's sequencing platform, polymerase enzymes are fixed on the bottom of the microcell, allowing only single DNA molecule to reach the DNA polymerase (Figure 1.5). The polymerisation is observed using 'Zero-mode waveguide' (ZMW) (Levene *et al.*, 2003). Once fluororescent dye-tagged nucleotide reaches the bottom of the well, the fixed DNA polymerase cleaves the dye, as the base is incorporated. The instrument determines the added nucleotide by detecting the laser-stimulating emission from cleaved fluorescent dye. The SMRT technology can produce sequence reads up to 10,000 bp or longer (Roberts, Carneiro & Schatz, 2013). The SMRT approach is also able to resolve GC rich regions. Moreover, this technology allows for the complete assembly of microbial genome (Ribeiro *et al.*, 2012). Nevertheless, SMRT sequencing can produce a noticeable indel error due to the single-molecule sequencing (Laehnemann, Borkhardt & McHardy, 2015). The utilisation of hierarchical genome assembly process (HGAP) (Chin *et al.*, 2013) and reads from other platforms (hybrid assembly) pave the way to the error correction of SMRT approach (Laehnemann, Borkhardt & McHardy, 2015).

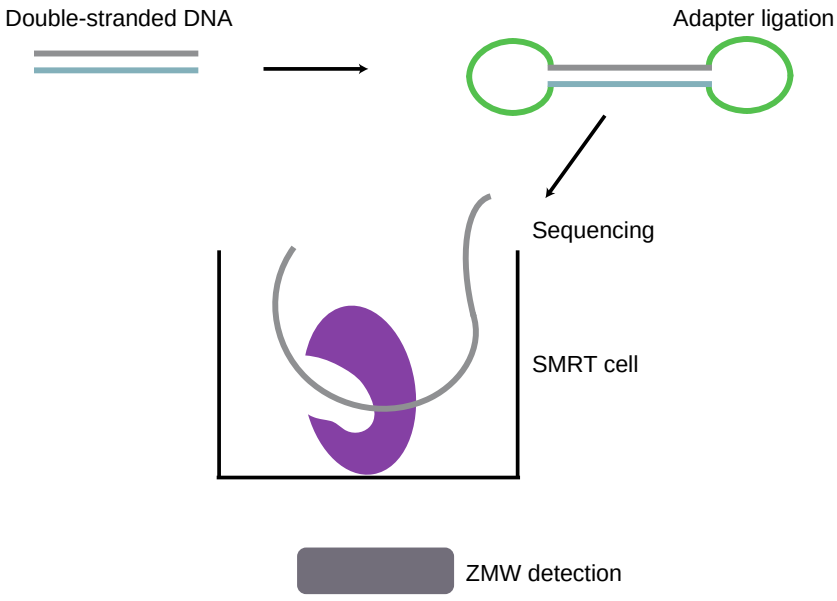


Figure 1.5: DNA sequencing by SMRT sequencing. Double stranded DNA template is ligated to adapter to achieve hairpin structure. Then the template is captured by bottom-fixed DNA polymerase and sequenced.

1.4.2. The application of whole genome sequencing in clinical microbiology

Advances in molecular biology and DNA sequencing have been improving biological researches. One implementation of DNA sequencing is to reveal bacterial genomes. The beginning era of bacterial genome dated back to 1995 when the first genome of *Haemophilus influenzae* Rd was successfully sequenced using the Sanger method (Fleischmann *et al.*, 1995).

Table 1.2: Examples of the application of whole genome sequencing approaches to bacterial infectious diseases.

Application	Study	Organisms	Reference
Identification and epidemiology	Outbreak in Germany	*Escherichia coli*	(Rasko *et al.*, 2011)
	Identification and typing	*Achromobacter spp.*	(Spilker, Vandamme & LiPuma, 2012)
	Typing, diversity and epidemiological study	Methicillin-resistant *Staphylococcus aureus**	(Enright *et al.*, 2000; Harris *et al.*, 2010, 2013)
	Outbreak in Spain	*Legionella spp.*	(Sánchez-Busó *et al.*, 2014)
	Population structure	*Haemophilus parasuis*	(Howell *et al.*, 2014)
	Outbreak in Denmark	*Salmonella spp.*	(Leekitcharoenphon *et al.*, 2014)
	Outbreak in ICU in Germany	*Klebsiella pneumoniae**	(Haller *et al.*, 2015)
Resistance gene identification	Tetracycline resistance	*Bifidobacterium animalis*	(Gueimonde *et al.*, 2010)
	Rifampicin resistance	*Mycobacterium tuberculosis*	(Comas *et al.*, 2012)
	Resistance integron	*Acinetobacter baumannii**	(Zhu *et al.*, 2014)
	Resistance prediction	*Escherichia coli*	(Tyson *et al.*, 2015)
	Resistome identification	*Pseudomonas aeruginosa*	(Kos *et al.*, 2015)
Virulence gene identification	Cobalamine biosynthesis and *eut* operon	*Morganella morganii**	(Chen *et al.*, 2012)
	Phosphilipase D family protein	*Klebsiella pneumoniae*	(Lery *et al.*, 2014)
	Virulence prediction	Methicillin-resistant *Staphylococcus aureus*	(Laabei *et al.*, 2014)
	A global transcriptional regulator of carbon catabolite repressor	*Enterococcus faecium*	(Somarajan *et al.*, 2014)
	Pathogenicity island prediction	*Helicobacter pylori*	(Ali *et al.*, 2015)

* hospital-acquired infections are included in the studies

The application of whole genome sequencing on microbiological studies has been continually reported as a routine use (Köser *et al.*, 2012; Fricke & Rasko, 2014), for example, diagnostic microbiology (Dunne, Westblade & Ford, 2012) and antibiotic resistance identification (Köser, Ellington & Peacock, 2014) (Table 1.2). This section aims to review the application of whole genome sequencing for microbiological studies, including, identification and gene profiling.

In diagnostic microbiology, precise identification methods are important. However, growing bacteria from clinical sample is a bottleneck. Some bacteria need a long time to grow or are uncultivable. Moreover, conventional identification using biochemical reaction-based approach cannot clearly identify species that have similar phenotypes. To improve bacterial species identification, the sequencing of 16S rDNA gene has been used on the basis of the DNA sequence of 16S rDNA is conserved within species (Janda & Abbott, 2007). Together with the increasing size of DNA sequence database, molecular diagnosis using gene sequencing has become a powerful approach to identify bacterial species. In addition, DNA sequencing can be used to classify bacteria into strain level. Maiden *et al.* (1998) proposed multilocus sequence-based analysis to discriminate bacterial species in strain level. Multilocus gene analysis uses multiple housekeeping genes, which are usually conserved within a species, to group the isolates into 'sequence types'. Multiloci-based analyses are currently used in many bacterial analyses, for example, *S. aureus* (Enright *et al.*, 2000; Feil *et al.*, 2003) and *Achromobacter spp.* (Spilker, Vandamme & LiPuma, 2012). Furthermore, whole genome sequencing has become an obvious implementation to investigate the epidemiology of species, in particular during the outbreak. Epidemiology of the species can also be determined based on the phylogenetic relationship (Ypma, van Ballegooijen & Wallinga, 2013). For example, an outbreak of haemolytic uraemic syndrome-causing *E. coli* in Germany was determined using the whole genome sequencing and comparative genome analysis, revealing horizontal gene exchange as a driving factor of the outbreak (Rasko *et al.*, 2011). Also, an outbreak of *Klebsiella pneumoniae* in neonatal intensive care unit was elucidated by the efficiency of whole genome sequencing (Haller *et al.*, 2015).

The whole genome sequencing is also applied to detect genes associated with antibiotic resistance in bacteria. An observation by Gordon *et al.* (2014) demonstrated that genome inspection was able to determine a correlation between known resistance genes and less susceptible phenotypes with the sensitivity and the specificity of 97% and 95%, respectively. This provides genetic information of antibiotic resistance genes for drug of choice for empirical treatment. Previous studies have demonstrated the application of whole genome sequencing to identify antibiotic resistance genes, for instance, the identification of tetracycline-resistance genes in *Bifidobacterium animalis* (Gueimonde *et al.*, 2010). Analyses using whole genome sequencing can address the dissemination of antibiotic resistance-associated moble genetic elements. This exemplifies by the identification of class 1 integron in *A. baumannii* using whole genome sequencing (Zhu *et al.*, 2014).

Apart from profiling antibiotic resistance genes, whole genome sequencing also has a potential to identify genetic determinants for the virulence of bacteria. For example, whole genome sequencing revealed genes encoding virulence-associated factors such as type 3 secretion system components and fimbrial adhesion protein in a human opportunistic pathogen *Morganella morganii* (Chen *et al.*, 2012). Also, whole genome sequencing, together with comparative genomic analysis, can predict virulence factors in pathogens, as exemplified by using a genome-wide association study for the prediction of virulence factors in MRSA (Laabei *et al.*, 2014)

1.4.3. The application of whole genome sequencing to emerging pathogens

Emerging pathogenic bacteria often lack information important for clinical investigation. This information includes potential virulence factors and antibiotic resistance genes. The application of whole genome sequencing for species identification and gene identification can be used for the study of emerging pathogens

This transformative technology has been enhancing an ability to deal with public health's problem. The identification of Shiga toxin 2-carrying phage leading to the outbreak of *E. coli* in 2009 (Rasko *et al.*, 2011) is an example of using whole genome sequencing to address the causing factor of the outbreak. In comparison with conventional microbiological study, whole genome sequencing does provide not only species identification but also genomic information of the pathogens. Researchers can retrieve a wealth of information, such as: species identity, virulence genes, antibiotic resistance genes, and genes associated with metabolic pathways, from the genome sequence. For example, a recent study addressed that the genome of *Campylobacter ureolyticus*, an emerging Gram-negative pathogen that is found associated with Crohn's disease in children, were approximately 9-20% conserved between *C. ureolyticus* and other *Campylobacter* species (Bullman *et al.*, 2013). Additionally, genomic analysis identified potential genes that are associated with the virulence of *C. ureolyticus.* Of 106 predicted virulence genes, 52 genes were predicted to be secretory proteins. It is believed that secretory proteins are the virulence factors in *C. ureolyticus* (Burgos-Portugal *et al.*, 2012). Genome sequence of *Mycobacterium abscessus* from clinical isolates was another successful application of whole genome sequencing to unravel emerging pathogen's genome (Ripoll *et al.*, 2009; Choo *et al.*, 2014). In addition to causing severe soft tissue infection, the multidrug resistant characteristics of *M. abscessus* has become a significant challenge (Nessar *et al.*, 2012). The genome sequence of *M. abscessus* revealed that phages were the driving components for the evolution of the species. Moreover, genome sequencing can be used to infer the pan-genome of *M. abscessus* was 'opened' (Choo *et al.*, 2014), referring to the possibility to receive external genes from the environment. With these examples of whole genome sequencing, this can be one of the approaches to the study of emerging pathogens. A combination of approaches, including whole genome sequencing, phenotypic test, susceptibility profile and virulence profile, is required for the establishment for linking genomic to biology.

1.5. *Achromobacter xylosoxidans* as an emerging pathogen

Achromobacter xylosoxidans was first described by (Yabuuchi & Oyama, 1971) as a Gram-negative bacteria presenting peritrichous pattern of flagella. Taxonomically, *A. xylosoxidans* is a member of the family *Alcaligenaceae* in the class β-proteobacteria. *A. xylosoxidans* has a close evolutionary relationship with *Bordetella*, which belong to the same family. When firstly described, *A. xylosoxidans* was only a member of the genus (Yabuuchi & Oyama, 1971). Until the present, the Genus is comprised of 19 species, including *A. aegrifaciens, A. animicus, A. anxifer, A. arsenitoxydans, A. cycloclates, A. denitrificans, A. dolens, A. insolitus, A. insuavis, A. lysyl, A. marplatensis, A. mucicolens, A. piechaudii, A. pulmonis, A. ruhlandii, A. sediminum, A. spanius, A. spiritinus,* and *A. xylosoxidans*. Nevertheless, *A. xylosoxidans* has the highest number of reports, clinically associated ones in particular, amongst the members of the Genus.

A. xylosoxidans can be found in various environments including soil, hotspring, municipal waste (Zhang *et al.*, 2007), and oil-polluted sea (Dave *et al.*, 2013). However, the bacterium can be seen as clinical samples. The first report of clinical isolates of *A. xylosoxidans* was from the infection of the middle ear. The strain was isolated and characterised biochemically by Yabuuchi & Oyama (1971). Subsequent studies reported the collection of *A. xylosoxidans* from hospitalised patients and immune-compromised patients (Holmes, Snell & Lapage, 1977; Pien & Higa, 1978; Shigeta *et al.*, 1978; Igra-Siegman, Chmel & Cobbs, 1980; Reverdy *et al.*, 1984; Arroyo *et al.*, 1987; Reina *et al.*, 1988). A major challenge of *A. xylosoxidans* infection is the antibiotic resistance of the species. The first report of ear discharge isolates showed that they were resistant to penicillin, ampicillin, erythromycin, gentamycin, kanamycin, streptomycin and tetracycline (Yabuuchi & Oyama, 1971). Later, Bador *et al.* (2011, 2013) characterised intrinsic efflux pumps associated with antibiotic resistance. Moreover, a gene coding oxacillinase was identified as an intrinsic gene and used as an species identification test (Turton *et al.*, 2011). The events of mobile genetic element in *A. xylosoxidans* have played a role in the acquisition of antibiotic resistance genes in the species, which worsened hospital-acquired

infections as the circulating pathogens can become drug-resistant strains (Shin *et al.*, 2005; Sofianou *et al.*, 2005; Yamamoto *et al.*, 2012; Traglia *et al.*, 2012). As a hospital-acquired infection-associated pathogen, several articles reported the outbreak of *A. xylosoxidans* in the hospital; for example, Reina *et al.* (1988) reported that contrast solution for computer tomography was contaminated by *A. xylosoxidans*. Tena *et al.* (2005) reported the persistent colonisation of *A. xylosoxidans* in intravascular catheter. Moreover, *A. xylosoxidans* has been considered an emerging pathogen in Cystic Fibrosis infection (Mahenthiralingam, 2014).

1.6. *A. xylosoxidans* as a Cystic Fibrosis pathogen

Cystic Fibrosis (CF) is a disease caused by mutation of an ion channel called 'Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)'. The mutation of the channel leads to the inability to secrete Chloride ion, and then mucus in the lumen becomes thicker and difficult to be expelled (Davies & Rubin, 2007). The mutation of the channel affects wide range of organs, such as respiratory tract, sweat gland, liver, intestine, and pancreas. Multiple clinical manifestations, such as, male infertility, intestinal obstruction, diabetes mellitus and obstructive pulmonary disease, are shown (Cutting, 2014). CF lung infection is a hallmark of cause of death (O'Sullivan & Freedman, 2009). The remained mucus in respiratory tract leads to serious clinical problems. The thickened mucus narrows the lumen of the tract, leading to airflow reduction during breathing. The unremoved mucus provides nourishing condition for bacterial colonisation *P. aeruginosa* has predominated CF infection; however, the prevalence of *P. aeruginosa* infection in CF decreased in the last decade (Lipuma, 2010). Other human pathogens, such as *S. aureus* and *Burkholderia cepacia* complex, also contribute to infesction in CF patients (Lipuma, 2010). Uncommon bacterial CF infection raises a challenge because less information is available.

A. xylosoxidans is considered as an emerging pathogen in CF (Mahenthiralingam, 2014). The prevalence of *A. xylosoxidans* infection ranges widely from 5% to 30% of CF infections depending on individual CF centre (Raso *et al.*, 2008;

Coburn *et al.*, 2015). Transmission of *A. xylosoxidans* from environment to CF patients remains unclear. Nevertheless, person-to-person transmission has been reported from several CF centres (Raso *et al.*, 2008; Magni *et al.*, 2010; Amoureux *et al.*, 2012; Hansen *et al.*, 2013). The major impacts of *A. xylosoxidans* infections are the severity of infection and antibiotic resistance of the species. Many studies showed that *A. xylosoxidans* infection worsened lung function of patients (De Baets *et al.*, 2007; Hansen *et al.*, 2010).

1.7. Complete genome sequence of the *A. xylosoxidans* genome

Because next-generation sequencing technologies have rapidly improved over the last decade, they are faster, cheaper and more accessible (van Dijk *et al.*, 2014). Complete genome sequences reveal large amounts of interesting and important information about bacteria, such as metabolic pathways, resistance genes and virulence factors. The first complete genome of *A. xylosoxidans* strain A8 was published in 2011 (Strnad *et al.*, 2011); however, MLST analysis shows that strain A8 is not placed in the *A. xylosoxidans* group (Spilker, Vandamme & LiPuma, 2012). The first complete genome of an *A. xylosoxidans* clinical isolate NH44784-1996 was composed of a total of 6,916,670 bp with 6,390 ORFs and 67% of GC content (Jakobsen *et al.*, 2013). The annotation revealed genes involved in anaerobic growth and biofilm formation. Moreover, as well as antibiotic-modifying enzymes, efflux transport systems were discovered in the genome and are expected to play an important role in major drug resistance mechanisms (Jakobsen *et al.*, 2013; Bador *et al.*, 2011, 2013).

1.8. Aims and objectives

At the beginning of this project, there was a need for effective and comprehensive analyses for the understanding of *A. xylosoxidans*. This work aims to compare methods for *A. xylosoxidans* identification and use comparative genome to analyse *A. xylosoxidans*. The comparative genomic study includes general genomic feature, antibiotic resistance, and virulence of the species.

1.8.1. The identification and strain typing of *A. xylosoxidans* (Chapter 3)

Many identification methods indicate the differentiation of genus *Achromobacter* from other bacterial genus, including *Burkholderia*, *Pseudomonas*, and *Bordetella*. A wide range of procedures, from conventional biochemical approaches to molecular-based approaches, have been applied with an effort to indicate the species. The aim of this study was to compare the efficiency of multiple identifications on clinical isolates of *A. xylosoxidans*, and then, based on a current analysis, an appropriate algorithm for *A. xylosoxidans* identification was proposed.

1.8.2. General feature of pan-genomic analysis of *A. xylosoxidans* (Chapter 4)

The advances in whole genome sequencing technology allow researchers to discover the genetic component of bacterial genome. The aims of this study were to investigate the genomic characteristic and functional genome of *A. xylosoxidans* and to investigate the diversity of the clinical isolates of *A. xylosoxidans* obtained from the U.K. and Thailand. Twenty-five genomes of *A. xylosoxidans* isolates collected in this study were therefore sequenced using next-generation sequencing. Genomic data was extracted to conduct comparative genomic analysis.

1.8.3. Comparative analysis of antibiotic resistance in *A. xylosoxidans* (Chapter 5)

Along with species identification, antibiotic resistance is another challenge that causes problems in clinical microbiology. This study aimed to extensive investigate genetic basis of antibiotic resistance in *A. xylosoxidans*. Class prediction analysis, along with antibiotic susceptibility profile, was applied to determine genes responsible for antibiotic resistance. Moreover, SMRT sequencing was used to investigate the presence of drug resistance-associated mobile genetic elements in the multidrug-resistant isolates.

1.8.4. The identification of virulence genes candidate in *A. xylosoxidans* (Chapter 6)

Infection with *A. xylosoxidans* causes a broad range of severity in patients who have immune-compromised conditions. A study in animal model demonstrated detrimental outcome of *A. xylosoxidans* infection in immune deficient mouse. This study aimed to determine the virulence of *A. xylosoxidans.* A wax moth larvae *Galleria mellonella* infection model was used to measure the virulence of the isolates. The prediction of virulence determinants was performed via class-predictive model to find the correlation between the presence of gene ortholgues and the virulence assessed by the infection model.

Chapter 2

General methodology and isolate description

2.1. Buffer for gel electrophoresis

2.1.1. Tris-Acetate-EDTA buffer (TAE buffer)

10X TAE buffer was prepared according to this following recipe

1 L of 10X TAE buffer

242g	Tris base
57.1ml	Glacial Acetic Acid
100ml	0.5M EDTA (pH8.0)

All ingredients were dissolved and made up to 1L using deionised water.

2.1.2. Tris-Borate-EDTA buffer (TBE buffer)

10X TBE buffer was prepared according to this following recipe

1 L of 10X TBE buffer

108g	Tris base
55g	Boric acid
140ml	0.5M EDTA (pH8.0)

All ingredients were dissolved and made up to 1L using deionised water.

2.2. Culture medium

One of most widely used culture medium is Lysogeny broth (LB). The standard ingredients of LB are Tryptone, Yeast Extract, Sodium Chloride and distilled water (Gerhardt, 1994). LB provides nutrients for the maintaining non-fastidious bacteria in culture. Although Blood Agar were used to recover *A. xylosoxidans* from clinical samples (Yabuuchi & Oyama, 1971; Gómez-Cerezo *et al.*, 2003; Tena *et al.*, 2005; Gomila *et al.*, 2014), haemolysis could not be detected in *A. xylosoxidans* (Otta *et al.*, 2014). Nutrient Agar/broth was therefore an appropriate medium to grow and maintain *A. xylosoxidans*. Consequently, LB was selected as the standard culture medium in liquid and agar form for *Achromobacter xylosoxidans* culture.

Recipe for 1L of LB (Gerhardt, 1994)

10g	Tryptone
5g	Yeast Extract
10g	Sodium Chloride
15g	Agar-agar (For agar only)

All ingredients were measured and dissolved in 1L of distilled water. The mixture was autoclaved at 121°C for 15 minutes.

2.3. Sample collection and culture

In this study, comparative genomic analysis of *A. xylosoxidans* was conducted on bacteria from three different sources; reference strains, clinical isolates from the U.K. and clinical isolates from Bangkok.

2.3.1. Reference strains

Six reference strains of bacteria in genus *Achromobacter* were purchased from Leibniz-Institut DSMZ, Germany and NCIMB, U.K. These strains were:

a) *Achromobacter xylosoxidans* DSM 2402, an ear discharge isolate (Yabuuchi & Oyama, 1971)
b) *Achromobacter denitrificans* DSM 30026, a soil isolate (Yabuuchi *et al.*, 1998)
c) *Achromobacter ruhlandii* DSM 653, a soil isolate (Yabuuchi *et al.*, 1998)
d) *Achromobacter piechaudii* DSM 10342, a pharyngeal swab isolate (Yabuuchi *et al.*, 1998)
e) *Achromobacter spanius* DSM 23806, a human blood isolate (Coenye *et al.*, 2003)
f) *Achromobacter insolitus* DSM 23807, a human wound isolate (Coenye *et al.*, 2003)
g) *Achromobacter xylosoxidans* NCIMB 11015, a soil isolate (Braker, Fesefeldt & Witzel, 1998)

These reference strains of *Achromobacter* have been well identified and used in several studies. These isolates were cultured on LB from the freeze-dried stocks, and cloned before freezing at -80^{o}C as glycerol stock aliquots.

2.3.2. Bangkok strain collection

Thai clinical isolates (Table 2.1) were collected from Ramathibodi hospital, a tertiary teaching hospital, in collaboration with Associate Professor Pitak Santanirand, Microbiology Unit, Department of Pathology, Faculty of Medicine, Ramathibodi hospital, Mahidol University. There were 14 isolates collecting between 2010 and 2012, see Table 2.1 for details. The collection of these samples was approved by the ethical review committee, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Thailand (MURA 2009/1396/S8; ID 03-52-93). After collection and identification, these strains were stored in Stuart

transfer medium and transferred to Institute of Infection and Global Health, University of Liverpool, Liverpool, U.K.

Table 2.1: A list of Thai clinical isolates and the detail of collection

Isolate code	Site of isolation	Type of specimen	Date received
R1	OPD (SURG)	Sputum	31.3.2012
R2	Ward (PED1)	Gastrostomy	22.1.2012
R3	ER	Wound (Pus)	10.3.2012
R4	ER	Urine	2.4.2012
R5	OPD (SURG)	Sputum (tube)	31.3.2012
R6	OPD (OBGYN)	Urine	19.3.2012
R7	ICU (SURG)	Blood (Peripheral)	2.10.2011
R8	ER	Blood (Peripheral)	21.12.2010
R9	Ward63 (SDMC)	Sputum	12.4.2012
R10	OPD (SURG)	Wound (Pus)	27.2.2012
R11	Ward (MED1)	BAL	15.3.2012
R12	OPD (SURG)	Right Buttock	19.1.2012
R13	ICU5 (SDMC)	Sputum	14.4.2012
R14	OPD (SURG)	Sputum	21.1.2012

OPD stands for out-patient clinic; ER for emergency room; ICU for intensive care unit; SURG for surgery unit; PED for paediatric unit; OBGYN for obstetric and gynaecology unit; MED for internal medicine unit; and SDMC for Somdej Bhradebaratana Medical Centre

2.3.3. British strain collection

There were 17 isolates collected from multiple sites, mainly in Liverpool, between 2008 and 2012, see Table 2.2 for details. These clinical isolates will be called 'Liverpool isolates' in this thesis. At the site of collection, they were identified using API20 NE kit before they were delivered to Institute of Infection and Global Health, University of Liverpool, Liverpool, U.K.

Table 2.2: A list of Liverpool clinical isolates and the detail of collection

Isolate code	Location	Specimen Type	Date Received	Clinical note
L1	Liverpool	Sputum	29.10.2011	COPD
L2	Liverpool	Blood culture	25.9.2010	Infected line
L3	Liverpool	Blood culture	29.7.2010	Catheterisation
L4	Liverpool	Blood culture	7.10.2010	Fever
L5	Liverpool	Blood culture	21.3.2009	Fever
L6	Liverpool	Blood culture	13.6.2008	Sepsis
L7	London	Eye discharge	25.7.2008	Eye infection
L8	Liverpool	Sputum	7.11.2011	COPD
L9	Liverpool	Blood culture	29.2.2012	CF
L10	Liverpool	Sputum	28.3.2012	Respiratory infection
L11	Liverpool	Sputum	28.3.2012	Bronchiectasis
L12	Liverpool	Node Tissue	18.4.2012	Lymphanenopathy
L13	Liverpool	No data	23.5.2012	NA
L14	Liverpool	Contact lens	23.5.2012	Keratitis
L15	Liverpool	Sputum	23.5.2012	Bronchiectasis
L16	Liverpool	Sputum	23.5.2012	COPD
L17	Liverpool	No data	20.5.2012	CF

COPD stands for Chronic Obstructive Pulmonary Disease; CF for Cystic Fibrosis; NA for not applicable

2.4. Culture and stock

Stock cultures were made by plating a single colony of each *Achromobacter* isolate in LB agar and incubating the plates at 37°C overnight. A single colony of cultures was harvested from overnight incubation and re-suspended in 25% glycerol in LB. The 25% glycerol LB suspensions were stored at -80°C ready to use in this study.

2.5. Growth of bacteria for experiments and genomic DNA

Bacteria where plated directly from frozen stocks onto LB agar plates and incubated at 37°C overnight. A single colony from these overnight cultures was inoculated into liquid LB and shaken at 37°C overnight (18 h). Mid-logarithmic growth phase cultures were obtained by inoculating an overnight liquid culture to fresh LB, 1:1000 dilution and shaking for 6 h at 37°C until 0.5-0.6 OD_{600} was reached. These mid-log cultures where used for the majority of experiments, including DNA extraction.

2.6. Genomic DNA purification and evaluation

2.6.1. DNA purification

The main objective of this project is to perform comparative genomic study of *A. xylosoxidans* from different sources, so the purification of genomic DNA from bacteria is the first step to disclosing the genetic information of *A. xylosoxidans*. There are a variety of DNA extraction methods used today, including phenol/chloroform, resin-based, and a membrane column methods. The phenol/chloroform preparation is widely used in molecular biology. However, both phenol and chloroform are corrosive chemicals, so handling with them should be done in a fume cupboard. According to health and safety, less harmful methods were used to purify DNA from bacteria in this project. Consequently, a resin-based purification was used for PCR reaction and a column-based purification was used for whole genome sequencing.

2.6.1.1. Single colony DNA extraction using 5% Chelex method

A 5% working stock was made by adding 1 g of Chelex®-100 (Bio-Rad) was added to 20 ml of sterile molecular grade water. After vigorous mixing the mixture was aliquoted and stored at 4°C prior to use (maximum one month). A single colony of each *A. xylosoxidans* isolate was collected from LB plate and re-suspended in a 5% Chelex® resin aliquot. After vortexing, samples were heated at 95 °C on a heat block for 5 min. The heated the mixture, was centrifuged at 13,000 x g for 2 min. The DNA was collected by transferring the supernatant to a new sterile 1.5-µL microtube and stored at 4 °C overnight.

2.6.1.2. Genomic DNA preparation using Membrane column-based preparation

Genomic DNA preparation followed the "Gram-negative bacteria" protocol provided as part of the DNeasy Blood & Tissue kit (QIAGEN, U.K.). In brief, a one ml mid-log culture of *A. xylosoxidans* was pelleted by centrifuging at 5,000 x g for 10 min. The supernatant was discarded and 180 µL of buffer ATL was added to resuspend cell pellet, followed by 20 µL of Proteinase K and mixed by vortexing. After incubation for 1 h at 56°C the clear suspension was mixed with 400 µL of pre-mixed buffer AL buffer, and transferred to a spin column provided and centrifuged for 1 min at 6,000 x g. Then washed with the AW1 and AW2 according to the manufactories instructions. Finally, DNA was eluted with 100 µL of sterile molecular grade water, incubated for 1 min prior to centrifugation at 6,000 x g for 1 min.

2.6.2. Quality control of DNA extraction

DNA for molecular biology and genome sequencing needs to be quality controlled for purity, concentration and molecular weight.

2.6.2.1. DNA Purity

The DNA extracted from an organism can be contaminated with a number of compounds, including RNA and protein. To validate the purity of DNA extractions, the spectrophotometric property of DNA, RNA and protein was measured to determine the contamination. Due to physical property of the substances, nucleic acids have a maximum absorbance at 260 nm and proteins have a maximum absorbance at 280 nm. Therefore, the ratio of absorbance at 260 nm to absorbance at 280 nm is applied to validate DNA preparation. Normally, an acceptable ratio is 1.8-2.0, because this ratio is equivalent to less than 60 % protein in the preparation. In this study, light absorbance of DNA extracts were assessed using a Nanodrop™ spectrophotometer. Molecular grade water was used a blank to set absorbance at 260/280 nm to zero. The optical density of DNA extract was measured for 3 μL of the extract.

2.6.2.2. Quantification of DNA concentration

In this study, we used the commercial fluorescent dye assay, Qubit® dsDNA BR kit (Life Technology), which covers a range from 2-1,000 ng of DNA following the manufactures instructions.

2.7. Whole Genome sequencing

Sequencing of genomic DNA of *Achormobacter* isolates was performed using next-generation sequencing machines at the Centre for Genomic Research (CGR), University of Liverpool. All *Achromobacter* isolates were sequenced on the Illumina HiSeq 2000 platform, and three *A. xylosoxidans* strains, NCIMB 11015, R4 and R8, were also sequenced using the single-molecule, real-time PacBio RSII platform. All the library preparation and sequencing was performed by the CGR.

Chapter 3

Comparison of methods for the identification and typing of *Achromobacter xylosoxidans*

3.1. Introduction

Methods for reliable species identification and strain typing are important for microbiological study, especially, clinical microbiology. Precise species identification of pathogens allows for the establishment of appropriate treatments and for the epidemiological study of the pathogens. However, the differentiation of bacterial species into strain level, called 'strain typing', improves the understanding of strains responsible for the outbreaks (Bosch *et al.*, 2010; Harris *et al.*, 2013). The introductory part of this chapter will present currently available methods for species identification and strain typing.

3.1.1. Identification of bacteria

The identification of bacteria is the keystone that shapes the understanding of microbiology. The identification methods used in the clinical and epidemiological study should be accurate, as to enable the suitable treatment and infection prevention strategies. Methods for bacterial identification have been developed over the time. Conventional culture and phenotype-based methods of determination are still widely used as screening tests even though molecular methods, such as PCR amplification and gene sequencing (for example of 16S rDNA), have shown better sensitivity and accuracy in bacterial identification over culture-based methods and biochemical reaction-based methods. For example, MacConkey agar has been used as selective agar for non-fastigious Gram-negative bacteria. The presence of crystal violet and bile salts in MacConkey agar inhibit the growth of most of Gram-positive bacteria and

fastidious Gram-negative bacteria (Adams, 1967). However, outer membrane of non-fastidious Gram-negative bacteria does not allow crystal violet and bile salts to get into the cell. Therefore, only non-fastidious Gram-negative bacteria can grow on MacConkey agar (MacConkey, 1900). Moreover, the agar has been used as agar for lactose fermentating bacteria differentiation due to the presence of lactose and pH indicator in the agar.

Biochemical analysis of bacteria is often able to identify the unique biochemical phenotypes associated with species of bacteria. However, such approaches are often limited in their ability to discriminate between strains of the same species. Therefore, bacterial identification has been improved by using molecular approaches: DNA-based methods and proteome-based methods.

3.1.1.1. Bacterial identification using the 16S rDNA gene

Since the advent of molecular techniques such as PCR amplification, molecular-based identification has been developed and widely introduced as an important tool to characterise bacterial species. In particular, the DNA sequence of 16S rDNA genes has been used extensively for bacterial identification and phylogenetic analyses to study relationships between bacterial species (Weisburg *et al.*, 1991).

Compared to conventional phenotypic tests, which can often vary upon repetition and might lack sensitivity and sensitivity, 16S rDNA sequencing is more reliable. 16S rDNA sequencing successfully identifies most common human pathogens which cannot be identified only by using conventional biochemistry (Woo *et al.*, 2001; Lau *et al.*, 2002; Woo *et al.*, 2008). In addition, PCR amplification coupled with 16S rDNA sequencing is able to find uncultivable pathogens of many important diseases, such as uncultivable *Tropheryma whippeli*, a causative pathogen of Whipple's disease (Relman *et al.*, 1992).

Some human infectious diseases are caused by emerging pathogens. The universal approach of 16S rDNA sequencing allows researchers not only to

diagnose causative known pathogens from clinical samples, but also to identify new species. The discovery of *Ehrlichia chaffeensis*, a causative pathogen in human ehrlichiosis exemplifies the application of 16S rDNA sequencing to identify emerging pathogens and discover new bacterial species because it is difficult to cultivate the pathogen from clinical samples (Anderson *et al.*, 1991).

3.1.1.2. Bacterial identification using MALDI-TOF mass spectrometry

Matrix-Assisted Laser Desorption Ionisation Time-Of-Flight mass spectrometry (MALDI-TOF MS) has been successfully proposed as an alternative bacterial identification method. This technique applies the principle of mass spectrometry whereby the mass-to-charge ratio (m/z) of proteins is measured, and then used to generate spectral fingerprints, which vary between the bacteria (Fenselau & Demirev, 1975). Before the analysis, the sample is pre-mixed with a matrix that helps the sample absorbing the light (Croxatto, Prod'hom & Greub, 2012). The sample can be either intact bacteria or bacterial protein extraction. The species identification is then performed by comparing those fingerprints to the database under certain growth conditions and sample preparation techniques. It is noteworthy to recognise that the spectral fingerprint can vary for the same isolate under different growing condition because different spectrum profiles can be affected by changes in growth conditions and growth medium (Carbonnelle *et al.*, 2011). Therefore, it is important to follow manufacturer's protocol, especially, type of culture media and incubation time. MALDI-TOF MS should be used with caution and as part of the evidence used for sample identification.

MALDI-TOF MS is becoming a routine identification method in clinical microbiology and a recommended method for the identification of unusual organisms (Carbonnelle *et al.*, 2011). In addition, MALDI-TOF MS has been reported to improve the resolution and the quality of bacterial identification from clinical isolates (Fernández-Olmos *et al.*, 2012). In comparison with other identification methods, mass spectrometry is claimed to be one of the most accurate and fast methods to diagnose the cause of infection (Schröttner *et al.*, 2014; Anderson *et al.*, 2014). It has been reported that MALDI-TOF MS showed

a significant increase in the correctness of species identification in Cystic Fibrosis infection pathogens (Degand *et al.*, 2008). In particular, the identification of non-fermenting Gram-negative bacteria from Cystic Fibrosis patients was improved. Nevertheless, MALDI-TOF MS cannot clearly identify uncommon or emerging pathogens because the limitation of the size of the spectral fingerprint database (Croxatto, Prod'hom & Greub, 2012; AbdulWahab *et al.*, 2015).

MALDI-TOF MS technology has provided clinicians and researchers with a number of advantages compared to biochemical identification methods. In terms of cost-effectiveness, using MALDI-TOF MS significantly reduced cost and waste disposal, compared to using biochemical identification (Gaillot *et al.*, 2011). Moreover, using MALDI-TOF MS decreased time for identification, compared to standard biochemistry (Tan *et al.*, 2012). Taken together, these studies suggest that MALDI-TOF is a better choice than biochemical identification, especially in the hospital setting.

3.1.2. Bacterial strain typing

Phenotypic and genotypic variations between bacteria can explain why bacteria with the same species behave differently. To define bacteria beyond the species level, especially for epidemiological benefit, strain typing is required. In the public health setting, bacterial strain typing can help clinicians to recognise particularly virulent strain types or strains that might be more difficult to counter; for example, *E. faecium* clonal complex 17 is associated with vancomycin resistance (Willems *et al.*, 2005). Additionally, strain type information provides useful data when an outbreak happens (Mora *et al.*, 2011), with the potential to unequivocally identify patient-to-patient spread and identify potential sources.

Many bacterial typing systems are based on molecular or genomic characteristics of the bacteria. There are 2 main categories of DNA-based methods for discrimination at the sub-species level: DNA fingerprinting and DNA sequencing. Traditional typing using fingerprinting, exploits variations in the relative

positions of particular sequences within the bacterial chromosome. For example, restriction enzyme digestion identifies variations in restriction sites and DNA amplification using arbitrary primers identifies variations in the relative positions of the sequences that the primers bind to.

3.1.2.1. Bacterial typing using DNA fingerprint

The most well-known restriction enzyme-based technique is pulsed field gel electrophoresis (PFGE), which involves the use of a rare-cutting restriction enzyme with subsequent band separation using a special electrophoretic technique (pulsed-field) to enable separation of large bands. The selection of restriction enzyme is the most essential step for PFGE (Goering, 2010). PFGE has high resolution and has been used for strain typing in many clinically important pathogens, such as *Salmonella* (Zou *et al.*, 2013), *S. aureus* (Bosch *et al.*, 2010), and *P. aeruginosa* (Fothergill *et al.*, 2010). Although, PFGE generates the power of strain discrimination and provides epidemiological information of the species, PFGE is difficult to compare between laboratories and is time-consuming.

Similar to the restriction enzyme-based techniques, fingerprinting using DNA amplification is on the basis of the variation of genomic DNA sequences between strains. Instead of using restriction enzymes, amplicons are generated by selected PCR primers which are short and therefore bind in multiple places. If two primers bind in the correct orientation in close enough proximity then an amplicon will be produced. The amplicons that occur are visualised on an agarose gel as a DNA fingerprint. Because the relative positions of primer-binding sites differ between strains, variations will occur in the banding patterns obtained. Random amplification of polymorphic DNA (RAPD) is an example of the DNA amplification–based methods. The term ‘random amplification of polymorphic DNA’ describes the random PCR products generated by short single oligonucleotide primers (< 10 bp) (Welsh & McClelland, 1990). In comparison to PFGE, RAPD does not require special electrophoresis; however, the resolution of RAPD is lower than that of PFGE (Kaur *et al.*, 2009). The

choice of PCR primers is an important factor for typing using RAPD. Similar to PFGE, the limitation of this technique is its low reproducibility between laboratories because variations occur readily if there are slight differences in protocols, such as types of buffer, primers and running time, or the thermal cycler used (Sabat *et al.*, 2013). Therefore, these fingerprint-based typing techniques can be used for a particular investigation. The standardisation and inter-laboratory reproducibility of fingerprint-based techniques is still necessary.

3.1.2.2. Bacterial typing using DNA sequences

DNA sequencing technologies have been developed and used by a large number of researchers. This typing is based on the similarity of DNA sequences and, consequently, the phylogenetic relationship of the strains. Typing using DNA sequence is considered more accurate than fingerprint-based methods because the sequence-based methods are using the comparison of DNA sequences, rather than DNA fragments or PCR amplicons (Li, Raoult & Fournier, 2009). The most commonly used DNA sequence-based typing method in genomic era is multilocus sequence typing (MLST) (Maiden *et al.*, 1998). This technique uses the DNA sequences of multiple housekeeping genes (usually seven genes) to generate a typing scheme. For each gene, each new sequence is given a defined allelic number. The relationship and the strain type are assigned by the combination of gene alleles. Since gene sequences can be stored in digital format, Jolley and his colleagues have developed an MLST database, which is available at PubMLST site (http://pubmlst.org). Consequently, this database establishes the standardisatio of MLST.

3.1.3. Identification and strain typing in *A. xylosoxidans*

One of the problems associated with emerging human pathogens is that often there is a lack of information about antibiotic resistance, virulence genes and diagnostic markers. As an emerging pathogen causing various types of infections in patients, identification methods for *A. xylosoxidans* are limited. Since the first report about *A. xylosoxidans* isolates from clinical samples, conventional

biochemical methods have been the standard way to identify *A. xylosoxidans* (Yabuuchi & Oyama, 1971). Liu *et al.* (2002) introduced the use of partial 16S rDNA sequence-targeted PCR assays to identify *A. xylosoxidans*. However, this approach is not specific to *A. xylosoxidans* and positive PCR results can occur in the closely related species *A. piechaudii*, *A. ruhlandii* and *A. denitrificans*. More recently, the intrinsic $Bla_{OXA-114}$ gene has been targeted as an alternative method for *A. xylosoxidans* identification. However, positive results were also present in *A. ruhlandii* (Turton *et al.*, 2011). MALDI-TOF MS has shown the improvement of identification of *A. xylosoxidans* (Degand *et al.*, 2008; Fernández-Olmos *et al.*, 2012), especially at the species level. Moreover, MALDI-TOF MS has demonstrated the ability to discriminate between *A. xylosoxidans* and other *Achromobacter* species (Gomila *et al.*, 2014).

Reliable typing methods for clinical isolates of *A. xylosoxidans* are required for epidemiological studies, in particular to monitor transmission of strains. Two of the most commonly used fingerprint-based typing strategies, PFGE and RAPD, have been applied to *Achromobacter*. Both PFGE and RAPD have been used to monitor the epidemiology of *A. xylosoxidans* (Amoureux *et al.*, 2012, 2013; Kaur *et al.*, 2009; Magni *et al.*, 2010; Krzewinski *et al.*, 2001). Typing using RAPD with primer 270 showed an association between RAPD profile of *A. xylosoxidans* and the severity of lung infection (decreased % FEV_1) by *A. xylosoxidans* (Magni *et al.*, 2010; Trancassini *et al.*, 2014). The association between PFGE typing and severity of *A. xylosoxidans* infection has not been reported. DNA sequence-based typing has, recently, been introduced as a more precise and reliable typing method for *Achromobacter*. Spilker, Vandamme & LiPuma (2012) applied MLST to generate strain typing for the genus *Achromobacter*. The genes used in the *Achromobacter* MLST analysis are *nusA*, *rpoB*, *eno*, *gltB*, *lepA*, *nuoL* and *nrdA*. Apart from typing the strain, MLST can be used to identify new species; for example, *A. xylosoxidans* AXX-A was re-named *A. insuavis*, based on MLST analysis (Spilker, Vandamme & LiPuma, 2012; Vandamme *et al.*, 2013)

3.1.4. Aims and objectives

There are increasing numbers of clinical isolates being identified as members of the genus *Achromobacter*, with *A. xylosoxidans* being the most abundant species detected. In addition, *A. xylosoxidans* causes more serious infection than the other *Achromobacter* species cause. However, the currently available identification procedures remain problematic for the identification of members of *Achromobacter*. Therefore, highly sensitive and specific methods for the identification of *A. xylosoxidans* are essential in order to ensure that clinicians choose the correct treatment options.

The aim of this study is to evaluate bacterial identification methods to find the most accurate and robust way to detect *A. xylosoxidans* amongst clinical isolates. A wide range of identification procedures will be compared: the determination of bacterial phenotype, 16S rDNA gene sequencing, and MALDI-TOF MS identification. Moreover, strain typing using RAPD-PCR and MLST are performed to study the relationships amongst the strains.

3.2. Material and methods

The cultures of *Achromobacter* were made by directly plating stock culture on LB agar and incubating at 37°C overnight. Plates were labelled with the code of isolates as described in section 2.3.

3.2.1. Biochemical property-based identification (Conventional method)

In collaboration with Ramathibodi Hospital in Bangkok, Thailand, all of the Thai clinical isolates were identified using traditional microbiological procedure at Ramathibodi Hospital, by courtesy of Dr Pitak Santaniand. For clinical isolates from Liverpool, all of them were previously identified using API20 NE kit prior to this study. To standardise the identification of the isolates, all British isolates was subjected to conventional identification.

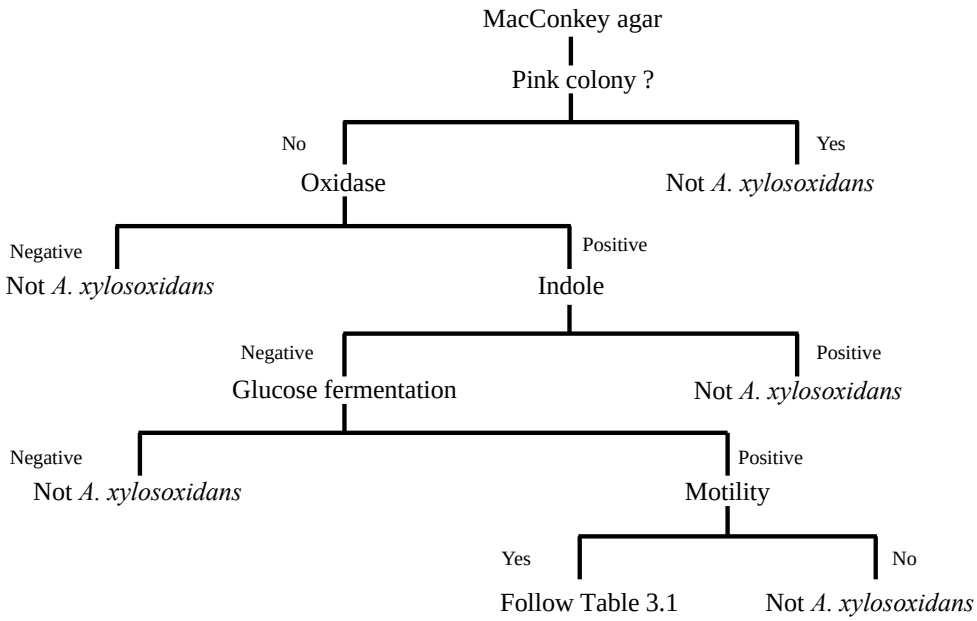


Figure 3.1: The identification algorithm for *A. xylosoxidans* (adapted from *Manual of Clinical Microbiology* (Murray *et al.*, 2007)

The identification of *A. xylosoxidans* was carried out as shown in Figure 3.1 with further biochemical analysis described in Table 3.1. All isolates were cultured on

MacConkey agar to investigate whether they were Gram-negative bacteria as well as to determine if they were lactose fermenters. According to Yabuuchi (Yabuuchi & Oyama, 1971; Yabuuchi & Yano, 1981; Yabuuchi *et al.*, 1998), *A. xylosoxidans* is a Gram-negative, motile, oxidase-positive, indole-negative, non lactose fermenting, glucose-fermenting, rod-shape bacterium with peritrichous flagellation.

Table 3.1: A summary of expected phenotypes of *A. xylosoxidans* based on biochemical properties. This table is a summary of table in *Manual of Clinical Microbiology* (Murray *et al.*, 2007).

Biochemical reaction	***A. xylosoxidans* (percent positive)**
Indole production*	0
Motility*	100
Sugar fermentation	
Lactose	0
Sucrose	0
Glucose	78
Xylose	99
Mannitol	0
Maltose	0
10% Glucose	100
Simmons' Citrate	95
Christensen's Urea	0
Nitrate reduction	100
Gas production from Nitrate	60
Nitrite reduction^{+}	0
Triple Sugar Iron (TSI) medium	
TSI slant, acid	0
TSI buttom, acid	0
TSI buttom, H2S	0
Protein / Amino acid utilization	
Gelatin hydrolysis	0
Aesculin hydrolysis	0
Lysine decarboxylation	0
Ornithine decarboxylation*	0

* Using Motility-Indole-Ornithine medium

$^{+}$ Nitrite reduction is performed in Nitrate reduction-negative and Gas production from Nitrate-negative isolates only

3.2.2. 16S rDNA sequencing

The diversity of the 16S rDNA gene was determined by species-specific PCR amplification 16S rDNA, using the primers shown in Table 3.2.

Table 3.2: 16S rDNA – based PCR primers used in this project (Liu *et al.*, 2002). AX-F1 and AX-B1 are specific to *Achromobacter*, whereas UFPL and URPL are universal primers for Kingdom *Bacteria*

Primer	Sequence (5' → 3')	Product size (bp)
AX-F1	GCAGGAAAGAAACGTCGCGGGT	163
AX-B1	ATTTCACATCTTTCTTTCCG	
UFPL	AGTTTGATCCTGGCTCAG	1,490
URPL	GGTTACCTTGTTACGACTT	

The PCR amplification was conducted in a total volume of 25 μL, containing 50 ng of genomic DNA template, 2.5 pmoles of primers, 1.25 mmol of $MgCl_2$, and 12.5 μL of 2X Biomix Red (Bioline, U.K.). Amplification was performed using a Veriti thermal cycler (Life Technology, USA) with the following conditions: 94°C for 1 min, followed by 35 cycles of 94°C for 1 min, 56°C for 45 s and 72°C for 1 min. Then, the mixture was heated to 72°C for 10 min in the final extension step. PCR products were visualised using 1% (w/v) agarose gel electrophoresis in TAE buffer (recipe described in section 2.1.1) at 100V for 45 min.

Amplicons generated using the UFPL and URPL primers were isolated using the ISOLATE PCR & Gel kit (Bioline, U.K.) and quantified by Qubit (Invitrogen, U.K.). Finally, the sequencing of PCR products was executed using Sanger sequencing (GATC Biotech, Germany). After obtaining the partial sequence of the 16S rDNA gene, the chromatograms of the sequence were manually verified via Geneious version 6.1.7 (Biomatters Ltd, New Zealand). After editing the sequences, nucleotide sequence alignments were processed using the MUSCLE

algorithm in the Geneious version 6.1.7 (Biomatters Ltd, New Zealand). A phylogenetic tree was then constructed by maximum likelihood method using PhyML (Guindon, 2010) with 1,000 bootstrap replications. The tree was visualised using Figtree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree). The partial sequences of 16S rDNA gene were used to search the database using BLAST (http://blast.ncbi.nlm.nih.gov). The option '16S ribosomal RNA sequences (Bacteria and Archeae)' and 'megablast' were selected for alignment database and for programme selection, respectively.

3.2.3. Matrix- assisted Laser Desorption / Ionisation – Time – of –Flight Mass spectrophotometry (MALDI-TOF MS)

A protein-based method to identify bacterial species was applied using MALDI-TOF mass spectrometry on a MALDI-TOF Biotyper (Broker Daltonic, Germany). A single colony was suspended in 200 µL of molecular grade water. Then 900 µL of absolute ethanol was added into the mixture and the mixture was centrifuged at 13,000 x g for 2 min. After supernatant removal, the mixture was centrifuged at 13,000 x g for 2 min. The residual ethanol was removed from the tube and the sample was set to air-dry. The pellet was mixed with 40 µL of 70% (v/v) formic acid. After leaving at room temperature for 2 minutes, 40 µL of absolute acetonitrile was added to the mixture, and then the mixture was centrifuged at 13,000 x g for 2 min. One microlitre of the supernatant was transferred to a steel plate and the sample was allowed to air-dry. Finally, 1 µL of α-Cyano-4-hydroxycinnamic acid matrix was overlaid on the air-dried sample, and the steel plate was inserted into the MALDI-TOF to analyse the sample. The analysis was conducted using built-in Biotyper software by comparing obtained spectra with a reference database provived. Species identification was achieved using an identification score ≥ 2.00.

3.2.4. Random amplification of polymorphic DNA (RAPD) PCR

Bacterial DNA was prepared by as previously described in section 2.6.1.1. and the concentration of DNA was adjusted to 40 ng/µL.

In this study, primer 270 (Mahenthiralingam *et al.*, 1996; Kaur *et al.*, 2009) was used to reveal discriminatory variation between isolates. Primer 270 is a 10-mer long DNA sequence: TGCGCGCGGG. To perform RAPD – PCR, the mixture was prepared in total volume of 25 µL, containing 40 ng of gDNA template, 7.5 pmol of primer 270 (Eurofin MWG), 1.25 µmol of $MgCl_2$ (Promega, USA), 0.5 µL of mixed dNTP (Promega, USA), 0.25 µL of GoTaq® DNA Polymerase (Promega, USA) and 2.5 µL of 10X Green GoTaq® flexi buffer (Promega, USA). PCR amplification was performed using Mastercycler® Thermal cycler (Eppendorf) using the following protocol, which was obtained from Kaur *et al.* (2009): 94°C for 15 minutes, then 4 cycles of 94°C for 5 minutes, 36°C for 5 minutes and 72°C for 5 minutes, followed by 30 cycles of 94°C for 1 minute, 36°C for 1 minute and 72°C for 1 minute, and 72°C for 10 minutes for the last step. The products were analysed using 1.5% (w/v) agarose gel electrophoresis in 0.5X TBE buffer using 95 Volts for 2 hours. Hyperladder I (Bioline) molecular weight marker was included on the gel. Finally, the DNA fingerprint from RAPD was analysed using Gelcompare II software (Applied Maths, Belgium).

3.2.5. Multi-locus sequence type (MLST)

Sequence typing of *A. xylosoxidans* was performed based on the standard seven housekeeping genes; *eno*, *gltB*, *lepA*, *nrdA*, *nuoL*, *nusA*, and *rpoB* (Spilker, Vandamme & LiPuma, 2012). The gene sequences of isolates were obtained from whole genome sequencing information (Chapter 4) and the gene sequences of other *Achromobacter* were obtained from the *Achromobacter* MLST database (http://pubmlst.org/achromobacter). After concatenating gene sequences, multiple sequence alignment was performed using Geneious version 6.1.7 (Biomatters Ltd, New Zealand) with MUSCLE and a maximum-likelihood phylogenetic tree was built on the alignment with 1,000 bootstrap replications. The analysis for population structure of *Achromobacter* isolates in this study was performed using eBURST algorithm (Feil *et al.*, 2004) implemented in goeBURST version1.2.1 (Francisco *et al.*, 2009).

3.3. Results

In this study, seven reference strains (described in section 2.3.1) have been used as species control to evaluate identification methods.

3.3.1. General phenotypic determination by conventional biochemical reactions

Phenotypic tests were carried out on seven reference strains (representing six different species), 17 clinical isolates from Liverpool and 14 clinical isolates from Thailand. According to Figure 3.1, all *Achromobacter* species are motile non-lactose fermenting bacteria. This was confirmed for all of the isolates in this study. With respect to saccharolytic ability, the type strains of *A. xylosoxidans*, NCIMB 11015 and DSM 2402, could use only glucose and xylose as carbohydrate carbon source, whereas non-*xylosoxidans* type strains could not use any sugar tested as a carbon source (Table 3.3). All clinical isolates, except L3, L4, L7, L12 and L13, could ferment glucose and xylose. All strains could grow on Simmons' citrate medium. *A. denitrificans*, *A. pichaudii*, *A. spanius*, *A. insolitus*, except L3, L4, L7, L12 and L13 gave positive results on 10% (w/v) glucose medium.

For Nitrogen utilising properties, most of clinical isolates including NCIMB 11015 and DSM 2402 could convert Nitrate into Nitrite and convert Nitrite into Nitrogen gas, which is detected by gas in a Durham tube (Table 3.3). Non-*xylosoxidans* type strains, L3, L4, L7 and L12 only converted Nitrate into Nitrite, which is detected by the appearance of red solution when alpha-naphthylamine and sulfanilic acid were added. The media of L13, L14 and L16 remained clear after the addition of alpha-naphthylamine and sulfanilic acid. However, with the addition of zinc powder, the solution with L13 became red, which means nitrate reduction is absence. The negative results of TSI agar and urea test indicated that all the isolates were aerobic bacteria with no expression of urease activity. The strains also showed inability to use Gelatin, Aesculin, Lysine and Ornithine as a carbon source (Table 3.3).

Considering the result of species identification, the phenotypic test assigned *A. xylosoxidans* to NCIMB 11015, DSM 2402, all Thai isolates and 12 British isolates. Some type strains were misidentified; *A. insolitus* and *A. spanius* were assigned *A. denitrificans*, and *A. piechaudii* and *A. ruhlandii* were assigned *Achromobacter sp.* (Table 3.3). Clinical isolates L3, L4, L7 and L12 were assigned *A. denitrificans*, and isolate L13 was assigned *Advenella incenata*, which is a relative of *Achromobacter sp* (Coenye *et al.*, 2005).

Table 3.3: Biochemical characteristic test on clinical *Achromobacter* isolates and *Achromobacter* type strains

Chemical tested	NCIMB 11015	DSM 2402	*A. denitrificans*	*A. ruhlandii*	*A. piechaudii*	*A. spanius*	*A. insolitus*
Motility	+	+	+	+	+	+	+
Sugar fermentation							
Lactose	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-
Glucose	+	+	-	+	-	-	-
Xylose	+	+	-	+	-	-	-
Mannitol	-	-	-	-	-	-	-
Maltose	-	-	-	-	-	-	-
10% glucose	+	+	-	+	-	-	-
Simmons' Citrate	+	+	+	+	+	+	+
Christensen's urea	-	-	-	-	-	-	-
Nitrate reduction	+[*]	+[*]	+[$]	+[$]	+[$]	+[$]	+[$]
Triple Sugar Iron (TSI)							
TSI slant, acid	-	-	-	-	-	-	-
TSI buttom, acid	-	-	-	-	-	-	-
TSI buttom, H2S	-	-	-	-	-	-	-
Protein utilisation							
Gelatin hydrolysis	-	-	-	-	-	-	-
Aesculin hydrolysis	-	-	-	-	-	-	-
Lysine decarboxylase	-	-	-	-	-	-	-
Ornithine decarboxylase	-	-	-	-	-	-	-
Result	*A. xylosoxidans*	*A. xylosoxidans*	*A. denitrificans*	*Achromobacter sp.*	*Achromobacter sp.*	*A. denitrificans*	*A. denitrificans*

[$] Positive when alpha-naphthylamine and sulfanilic acid added; [*] Nitrogen gas production from Nitrate reduction; [#] media remain clear when zinc powder added

Table 3.3: Biochemical characteristic test on clinical *Achromobacter* isolates and *Achromobacter* type strains (Continued)

Chemical tested	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17
Motility	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sugar fermentation																	
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glucose	+	+	-	-	+	+	-	+	+	+	+	-	-	+	+	+	+
Xylose	+	+	-	-	+	+	-	+	+	+	+	-	-	+	+	+	+
Mannitol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Maltose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10% glucose	+	+	-	-	+	+	-	+	+	+	+	-	-	+	+	+	+
Simmons' Citrate	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Christensen's urea	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nitrate reduction	+[*]	+[*]	+[$]	+[$]	+[*]	+[*]	+[$]	+[*]	+[*]	+[*]	+[*]	+[$]	-	#	+[*]	#	+[*]
Triple Sugar Iron (TSI)																	
TSI slant, acid	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSI buttom, acid	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSI buttom, H2S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Protein utilisation																	
Gelatin hydrolysis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aesculin hydrolysis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lysine decarboxylase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ornithine decarboxylase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Result	AX	AX	AD	AD	AX	AX	AD	AX	AX	AX	AX	AD	AC	AX	AX	AX	AX

[$] Positive when alpha-naphthylamine and sulfanilic acid added; [*] Nitrogen gas production from Nitrate reduction; [#] media remain clear when zinc powder is added

AX stands for *A. xylosoxidans*; AD for *A. denitrificans*; AC for *Advenella incenata*; and L1-L17 for British isolates 1-17

Table 3.3: Biochemical characteristic test on clinical *Achromobacter* isolates and *Achromobacter* type strains (Continued)

Chemical tested	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Motility	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sugar fermentation														
Lactose	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Xylose	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Mannitol	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Maltose	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10% glucose	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Simmons' Citrate	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Christensen's urea	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nitrate reduction	+*	+*	+*	+*	+*	+*	+*	+*	+*	+*	+*	+*	+*	+*
Triple Sugar Iron (TSI)														
TSI slant, acid	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSI buttom, acid	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSI buttom, H2S	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Protein utilisation														
Gelatin hydrolysis	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aesculin hydrolysis	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lysine decarboxylase	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ornithine decarboxylase	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Result	AX	AX	AX	AX	AX	AX	AX	AX	AX	AX	AX	AX	AX	AX

$ Positive when alpha-naphthylamine and sulfanilic acid added; * Nitrogen gas production from Nitrate reduction; # media remain clear when zinc powder is added

AX stands for *A. xylosoxidans*; and R1-14 for Thai isolates 1-14

Table 3.4: Summary of species identification using multiple approaches

Isolate	Biochemical properties	API20NE	16S rDNA PCR	16S rDNA sequencing	MALDI-TOF	MLST (ID)	MLST (ST)
NCIMB 11015	AX	N/A	Positive	AX	AX	AX	166
DSM 2402	AX	N/A	Positive	AX	AX	AX	20
AD	AD	N/A	Positive	AD	AD	AD	102
AI	AD	N/A	Positive	AI	AI	AI	99
AP	*Achromobacter sp.*	N/A	Positive	AP	AP	AP	122
AR	*Achromobacter sp.*	N/A	Positive	AR	AR	AR	30
AS	AD	N/A	Positive	AS	AS	AS	112
R1	AX	N/A	Positive	AX	AX	AX	27
R2	AX	N/A	Positive	AX	AX	AX	196
R3	AX	N/A	Positive	AX	AX	AX	182
R4	AX	N/A	Positive	AX	AX	AX	183
R5	AX	N/A	Positive	AX	AX	AX	27
R6	AX	N/A	Positive	AX	AX	AX	180
R7	AX	N/A	Positive	AX	AX	AX	184
R8	AX	N/A	Positive	AX	AX	AX	185
R9	AX	N/A	Positive	AX	AX	AX	186
R10	AX	N/A	Positive	AIn	AX	AX	182
R11	AX	N/A	Positive	AX	AX	AX	187
R12	AX	N/A	Positive	AX	AX	AX	188
R13	AX	N/A	Positive	AX	AX	AX	189
R14	AX	N/A	Positive	AX	AX	AX	27

N/A stands for Not applicable; AX for *A. xylosoxidans*; AD for *A. denitrificans*; AI for *A.* insolitus; AP for *A. piechaudii*; AR for *A. ruhlandii*; AS for *A. spanius*; AIn for *A. insuavis*; and R1-14 for Thai isolates 1-14

Table 3.4: Summary of species identification using multiple approaches (Continued)

Isolate	Biochemical properties	API20NE	16S rDNA PCR	16S rDNA sequencing	MALDI-ToF	MLST (ID)	MLST (ST)
L1	AX	AX	Positive	AX	AX	AX	28
L2	AX	AX	Positive	AIn	AX	AIn	168
L3	AD	AX	Positive	AX A8	*Achromobacter sp.*	ASp	169
L4	AD	AX	Positive	AP	AR	AP	170
L5	AX	AX	Positive	AX	AX	AX	171
L6	AX	AX	Positive	AX	AX	AX	28
L7	AD	AX	Positive	APu	AR	Genogroup19	172
L8	AX	AX	Positive	AX	AX	AX	175
L9	AX	AX	Positive	AX	AX	AM	174
L10	AX	AX	Positive	AX	AX	AX	173
L11	AX	AX	Positive	AX	AX	AX	176
L12	AD	AD	Positive	AM	Poor ID	ASp	177
L13	AC	AX	Positive	AS	AP	AS	178
L14	AX	AX	Positive	AX	AX	AX	179
L15	AX	AX	Positive	AX	AX	AX	176
L16	AX	AX	Positive	AX	AX	AX	180
L17	AX	AX	Positive	AX	AX	AX	175

AX stands for *A. xylosoxidans*; AD for *A. denitrificans*; AI for *A.* insolitus; AP for *A. piechaudii*; AR for *A. ruhlandii*; AS for *A. spanius*; AC for *Advonella incenata*; APu for *A. pulmonis*; AM for *A. marplatensis*; Asp for *A. spiritinus*; and L1-L17 for British isolates 1-17

3.3.2. Automated platform using API20NE and MALDI-TOF MS could not clearly identify *Achromobacter* species

Using the API20 NE identification database, sixteen clinical strains from Liverpool were identified as *A. xylosoxidans* and L12 was identified as *A. denitrificans* (Table 3.4). Bruker biotyper MALDI-TOF MS assigned all 14 Thai isolates and 12 British isolates to *A. xylosoxidans*. Opposing to the identification by API20 NE, MALDI-TOF MS reported that five British isolates were not *A. xylosoxidans*. L4 and L7 were *A. ruhlandii* and L13 was *A. piechaudii*. L3 was identified as *Achromobacter sp.* and L12 were reported as 'not reliable identification' by the Bruker biotyper database. The type strains of *Achromobacter* species were identified as their own type species using the MALDI-TOF MS Bruker biotyper database (Table 3.4). For Thai clinical strains, the MALDI-TOF MS assigned *A. xylosoxidans* to all isolates.

3.3.3. Genus-specific 16S ribosomal RNA-targeted identification

The results of amplification using primers AX-F1 and AX-B1 revealed the presence of a 163-bp-long amplicon from all type strains (Figure 3.2, Table 3.4). In the lane containing *P. aeruginosa* PAO1, there was no 163-bp long product but there were some DNA bands appearing on the top of the lane (PAO1, Figure 3.2). These bands did not exist on the other gels (Figure 3.3A), so the bands could be the degraded DNA or non-specific amplicons. PCR products were obtained from all clinical isolates and the positive control strain, NCIMB 11015 (Figure 3.3). Interestingly, the products were also detected from seven clinical isolates, L2, L3, L4, L7, L9, L12, and L13 (Table 3.4).

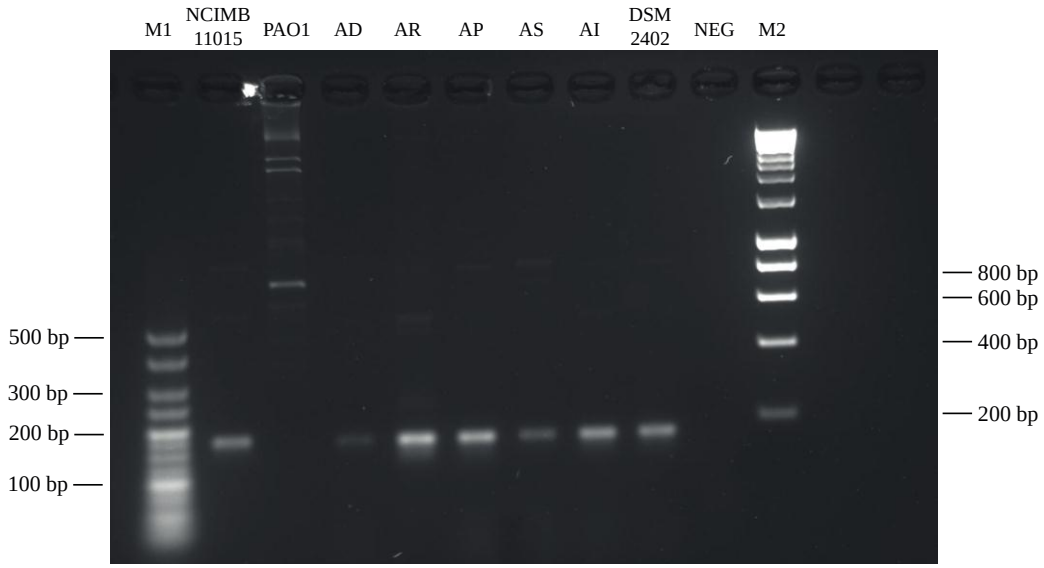


Figure 3.2: *A. xylosoxidans* "specific" 16S rDNA gene PCR applied to type strains of the genus *Achromobacter* used in this study. Label: M1 stands for DNA marker - Hyperladder V; NCIMB 11015 for *A. xylosoxidans* NCIMB 11015; PAO1 for *P. aeruginosa* PAO1; AD for *A. denitrificans*; AR for *A. ruhlandii*; AP for *A. piechaudii*; AS for *A. spanius*; AI for *A. insolitus*; DSM 2402 for *A. xylosoxidans* DSM 2402; NEG for Negative control; and M2 for DNA marker - Hyperladder I

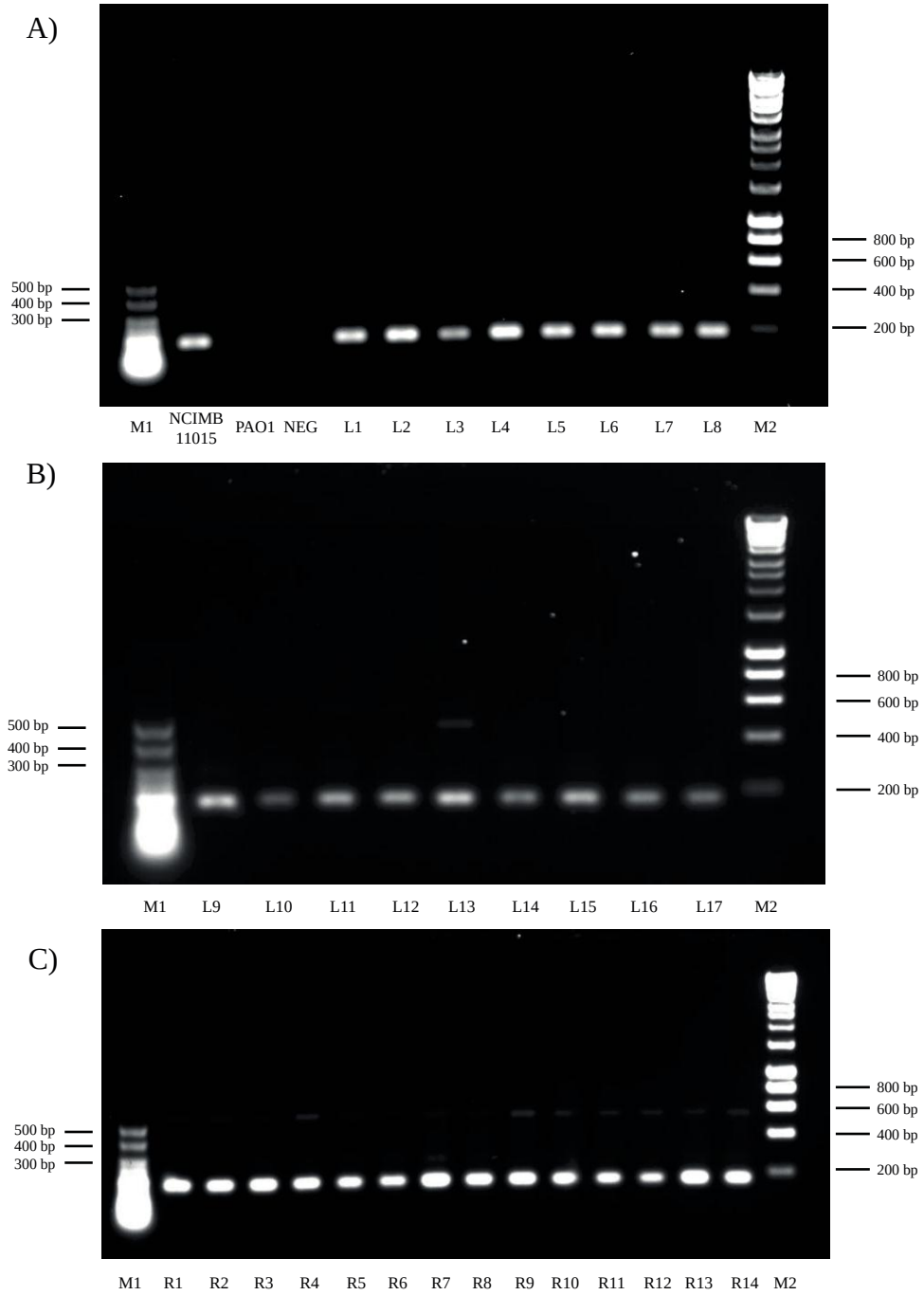


Figure 3.3: *A. xylosoxidans* "specific" 16S rDNA gene PCR applied to clinical *Achromobacter* isolates A) L1-L8, B) L9-L17 and C) R1-R14 using primers AX-F1 and AX-B1. Label: M1 stands for DNA marker - Hyperladder V; NCIMB 11015 for *A. xylosoxidans* NCIMB 11015; PA01 for *P. aeruginosa* PAO1; NEG for Negative control; M2 for DNA marker - Hyperladder I; L1-L17 for British isolates 1-17; and R1-R14 for Thai isolates 1-14.

3.3.4. Random amplification of polymorphic DNA patterns showed intra-hospital relationship of clinical isolates

A total number of 38 strains of *Achromobacter* were typed using RAPD-PCR. With primer 270, the profile suggests the colonisation of the same strain in the patients (Figure 3.4). For Thai isolates, the RAPD profile revealed the similarity of isolates R1, R5 and R14 (black boxes in Figure 3.4). This similarity was explained by the fact that these isolates were from the same patient. For isolates from Liverpool, L2, L3, L4, L12 and L13 displayed RAPD profiles that were different from the other strains (Figure 3.4). L11 and L15, which were taken from the same patient, showed similar RAPD pattern (Red boxes in Figure 3.4). These demonstrated persistent colonisation of *A. xylosoxidans*. In addition, the profile presented a pair as examples of colonisation of similar strain in different patients; L8 and L17.

With hierarchical clustering with UPGMA method in Gelcompare II, the analysis revealed non-*xylosoxidans* cluster (highlighted by orange box in Figure 3.5). This cluster consisted of *A. denitrificans*, *A. insolitus*, *A. spanius*, *A. ruhlandii*, L3, L12 and L13. However, some non-*xylosoxidans* strains and clinical isolates of *A. xylosoxidans* were clustered together, as indicated in a blue box in Figure 3.6. Considering *A. xylosoxidans* clusters, the dendrogram demonstrated a close genomic relationship (80% similarity) of isolates R1, R5 and R14 (Green box in Figure 3.5). Liverpool isolates that had similar RAPD pattern were also clustered together; L8 and L17 (Yellow box in Figure 3.5) were 80% similarity and L11 and L15 (Purple box in Figure 3.5) were 80% similarity.

3.3.5. Multilocus sequence typing analysis

The multilocus sequence analysis (Figure 3.6) presented better resolution of species identification, in comparison with 16S rDNA phylogeny (Figure A2.1). Here, with the MLST phylogenetic tree, strain L7 was grouped with *Achromobacter* genogroup 19, and strain L9 was grouped with *A. marplatensis*. Also, the MLST phylogeny assigned species/genogroups to non-*xylosoxidans* isolates; L2 was genogroup 2b, L3 and L12 were *A. spiritinus*, L13 was *A. spanius*, and L4 was *A. piechaudii* (Figure A2.2).

Considering a group of *A. xylosoxidans* on MLST phylogeny, some clinical isolates were assigned the same strain type; L8 and L17 were ST175, L11 and L15 were ST176, R1, R5 and R14 were ST27, and R3 and R10 were ST182 (Table 3.4). As shown by MLST analysis, clinical isolates of *A. xylosoxidans* used in this study were diverse (A black dashed box, Figure 3.6). A population snapshot using eBURST algorithm revealed 14 groups. The largest groups contained 18 STs with ST 180 as a founder of the groups (Appendix Figure A2.3). None of them had ‘single locus variant’ relationship.

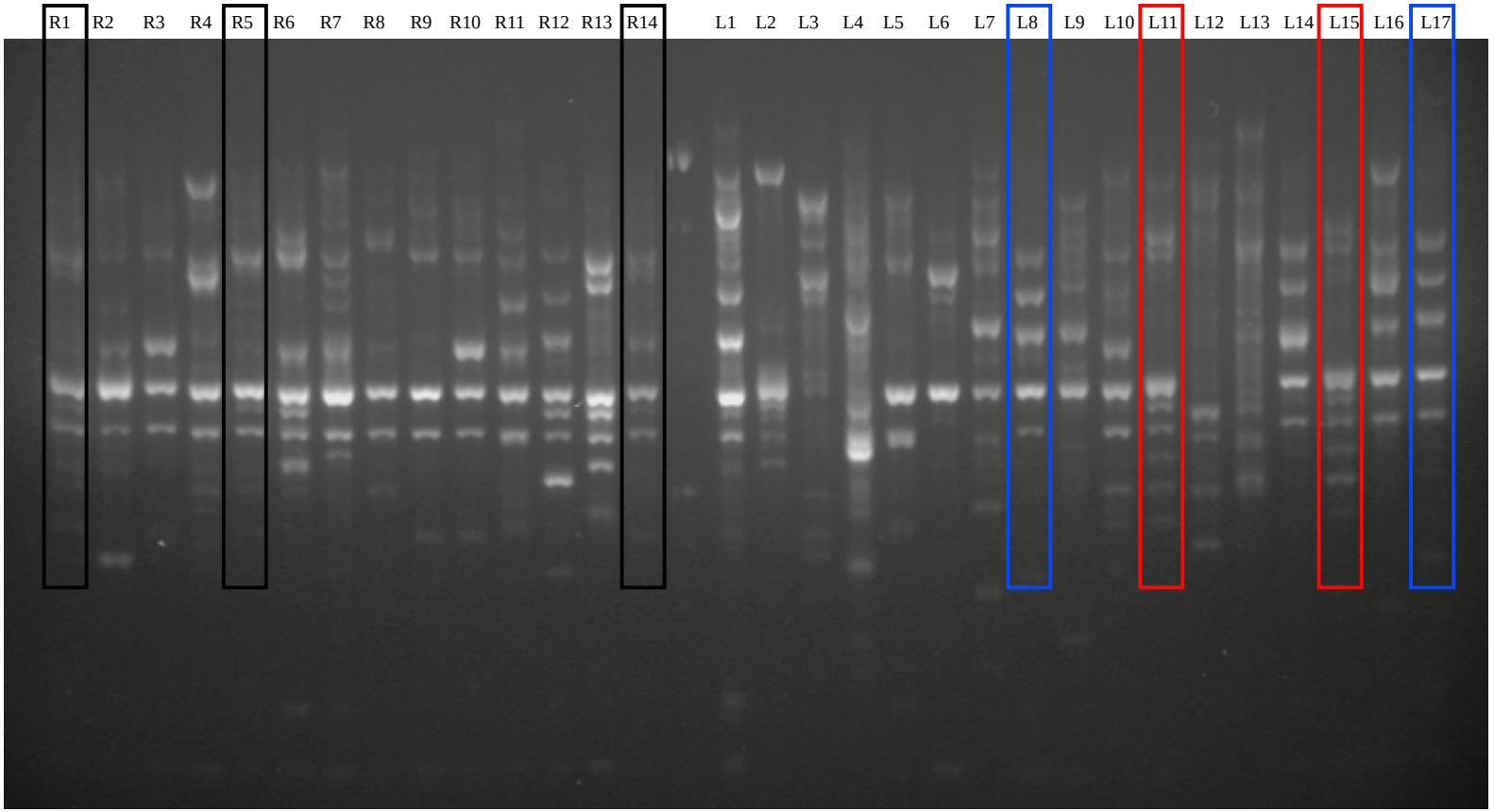


Figure 3.4: Gel electrophoresis showing the RAPD profile of 31 clinical isolates of *Achromobacter* species. The typing is performed using primer 270. Each box indicated the similarity of RAPD profile by inspection. R1-R14 stands for Thai isolates; and L1-L17 for British isolates

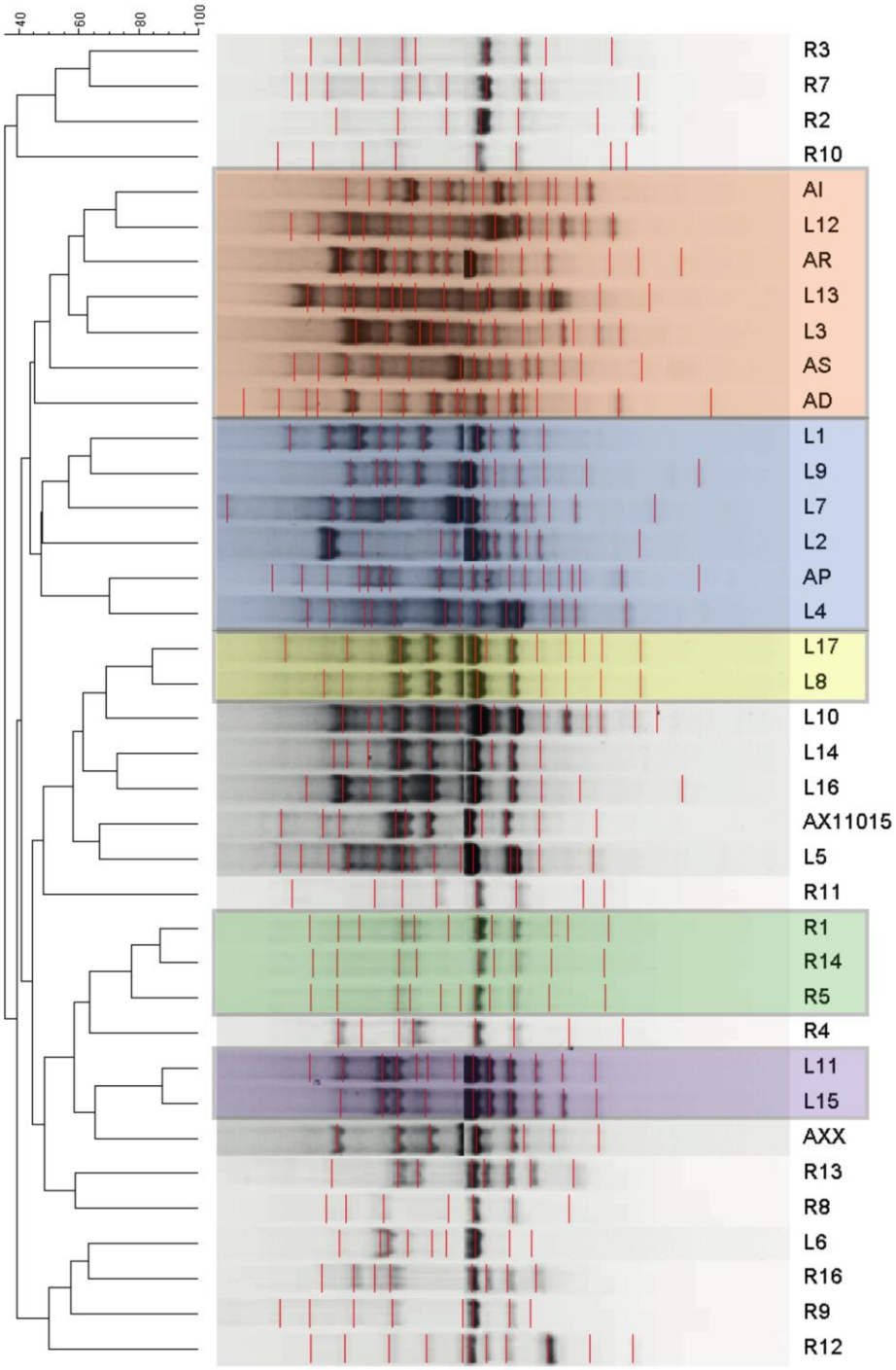


Figure 3.5: RAPD analysis of each *Achromobacter* isolates using primer 270. A dendrogram is constructed using UPGMA in Gelcompar II software. Label: R1-R14 stands for Thai isolates; L1-L17 for British isolates; AD for *A. denitrificans*; AR for *A. ruhlandii*; AP for *A. piechaudii*; AS for *A. spanius*; AI for *A. insolitus*; AXX for *A. xylosoxidans* DSM 2402; and AX11015 for *A. xylosoxidans* NCIMB 11015

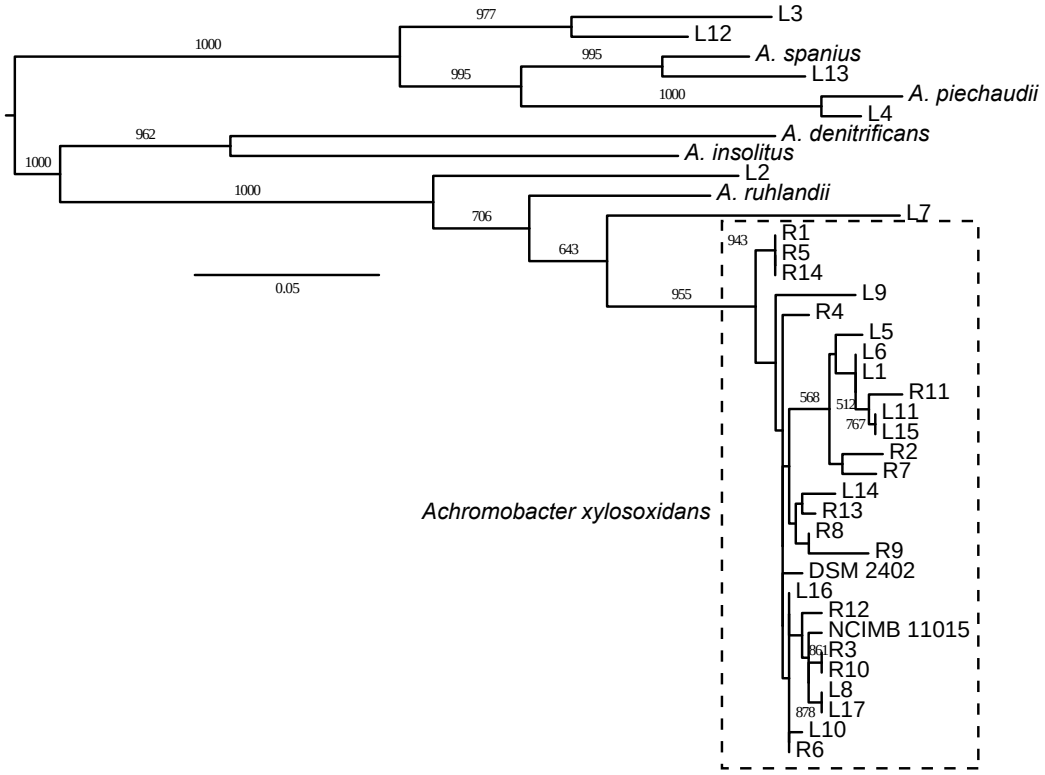


Figure 3.6: MLST phylogenetic tree of all *Achromobacter* isolates based on seven housekeeping genes. R1-R14 are clinical isolates from Thailand; L1-L17 are clinical isolates from the U.K.; DSM 2402 stands for *A. xylosoxidans* DSM 2402; and AX11015 for *A. xylosoxidans* NCIMB 11015. Bootstrap values >50% are indicated.

3.4. Discussion

3.4.1. Comparison of species identification methods

Precise identification methods for *A. xylosoxidans* are essential for clinical microbiology due to epidemiological study and the increase in the prevalence of clinical isolates being reported as *A. xylosoxidans*. Here, multiple identification approaches, ranging from conventional phenotypic techniques to molecular techniques, were applied to reveal appropriate and robust identification methods.

'A gold standard' of this study is determined by the ability to identify correct species from reference strains and by reviewing literatures. In the first place, the methods used to identify *A. xylosoxidans* were based on phenotypic or biochemical tests (Yabuuchi & Oyama, 1971). However, those methods were not robust enough for hospital's practice because the tests of bacterial phenotype need at least 24 hours for bacterial culture and a large panel of biochemical reaction. Therefore, molecular biology-based methods were taken into consideration. Of applied methods in this study, methods that correctly assign correct species to the reference strains are 16S rDNA sequence, MALDI-TOF MS, and MLST. Considering the identification of clinical isolates, 16S rDNA sequence shows the limitation to identify ambiguous clinical isolates. A recent study reveals that MALDI-TOF MS and MLST are methods that are able to identify and discriminate *Achromobacter* species in species and strain level. However, MALDI-TOF MS is limited to identify some clinical isolates in this study because of the small size of database. Therefore, MLST is considered a 'gold standard' approach in this study.

The comparison of biochemical phenotype-based approaches showed that manual biochemical reaction determination was better than API20 NE kit in terms of identification performance (Table 3.4). A previous study (Fernández-Olmos *et al.*, 2012) showed that API20 NE identified ten false positive *A. xylosoxidans* out of 66 strains (84.8% sensitivity), with respect to MLST identification. It also identified 25 false positive strains out of 59 strains as

Burkholderia cepacia complex (50.7% sensitivity). Here, in this study, with respect to identification by MLST, these approaches assigned the genus *Achromobacter* to all isolates. Taking the British isolates only, the phenotype-based tests misidentified seven strains out of 17 strains (58.8% sensitivity). In order to identify *A. xylosoxidans*, the classic biochemical test identified five out of seven non-*xylosoxidans* Liverpool isolates (71.4% sensitivity), whereas the API20 NE could only identify one out of seven non-*xylosoxidans* Liverpool isolates (14.3% selectivity). The limitation of species identification by biochemical reaction-based approaches can be explained by the restricted number of biochemical reactions that allow for better differentiation *Achromobacter* species. Therefore, the approaches could be appropriate as an initial screening test, but identification at the species level would require the use of additional methods.

Contrary to biochemical reaction-based approaches, the MALDI-TOF MS platform and 16S rDNA sequencing showed better performance on species identification. As shown in Table 3.4, these methods were able to differentiate *A. xylosoxidans* from other *Achromobacter* species. For five British isolates, L3, L4 L7, L12, and L13, all three identification methods were in agreement that they were not *A. xylosoxidans*. Nonetheless, the identification for certain species is still ambiguous, as illustrated by inconsistent identification of seven British clinical isolates (Table 3.4). The ambiguity of *Achromobacter* identification using 16S rDNA sequence was also reported (Gomila *et al.*, 2014). Considering available identification methods that have been used in hospital, MALDI-TOF MS has exhibited better performance to discriminate *Achromobacter* species with better sensitivity, compared to biochemical approach and 16S rDNA sequencing approach. Nevertheless, MALDI-TOF MS analyses are observed to be less accurate than MLST. On account of working by matching the queried spectra to the database, this suggests that the size of the database is essential for MALDI-TOF MS (van Veen, Claas & Kuijper, 2010).

As presented in this study, a combination of methods is required for accurate *A. xylosoxidans* identification (Gomila *et al.*, 2014). Biochemical reaction-based panel were reliable enough for genus level, but not for species level, so the panel

can be used as a screening test prior to executing further identification. MALDI-TOF, 16S rDNA sequencing and MLST demonstrated trustworthy identification. In clinical microbiology, a faster approach is preferred because early treatment to stop disease progression is important. In this respect, MALDI-TOF is better than the other methods because it is less time consuming. However, 16S rDNA sequencing and MLST identification could be supporting methods for strain-level identification.

3.4.2. Strain typing of *Achromobacter* members to study the global population

In comparison with many bacterial pathogens, the global population of *Achromobacter* has been poorly characterised. Several studies have reported the variation of *A. xylosoxidans* isolates and suggested that bacteria could spread out in various environments (Amoureux *et al.*, 2012, 2013). This study focused on applied multiple typing approaches to help discriminate *A. xylosoxidans* from other *Achromobacter* and to investigate the clustering of *A. xylosoxidans* from different geographical background.

The combination of DNA fingerprint typing and multilocus gene typing was applied to investigate species segregation and the epidemiology of *A. xylosoxidans*. In this study, RAPD fingerprinting and MLST were selected as strain typing methods because these methods are generally used in this species (Kaur *et al.*, 2009; Magni *et al.*, 2010; Spilker, Vandamme & LiPuma, 2012; Vandamme *et al.*, 2013; Trancassini *et al.*, 2014). Both RAPD fingerprinting and MLST analysis illustrated the ability to differentiate between *A. xylosoxidans* and non-*xylosoxidans Achromobacter* species, except that L1 was grouped with non-*xylosoxidans* by RAPD (Figure 3.5). Considering only *A. xylosoxidans*, both typing methods could not group *A. xylosoxidans* isolates by their countries of origin (Figure 3.5 and Figure 3.6).

By observing epidemiology of the isolates, the presence of similar genotypes of strains in different patients was observed in this study. Nevertheless, both the

RAPD typing and MLST demonstrated multiple genotypes of *A. xylosoxidans*, especially Thai strains (Figure 3.4). Both RAPD and MLST demonstrated multiple strain types of clinical isolates and the strains from both British and Thailand were not separated by geography.

As shown in Figure 3.5 and Figure 3.6, there were two pairs of genetically similar strains; L11-L15 and L8-L17. For Thai strains, clonal strains were unsurprisingly found in strain R1, R5 and R14. British strain L11 and L15 were obtained from the same patient (two months apart), who had bronchiectasis, a disease with abnormally widened lung bronchi. Thai strain R1, R5 and R14 were also collected from the same patient (two months apart) who had pneumonia. These evidences represented long-termed colonisation of *A. xylosoxidans,* at least, over two months in a single patient with respiratory tract problems during intubation (Table 2.1). Interestingly, L8 and L17 were isolated from a non-Cystic Fibrosis patient with chronic obstructive pulmonary disease and a Cystic Fibrosis patient, respectively. It is still argued that how those patients be infected by similar strain type. It can be assumed that these patients contacted with the same environment so that they were infected with genetically-related strain of *A. xylosoxidans*.

The diversity of *A. xylosoxidans* isolates demonstrates that the majority of infected patients acquire the pathogens from the environment. There is a possibility of persistent colonisation in patients with respiratory diseases (Kanellopoulou *et al.*, 2004). Furthermore, these evidences suggest no geographic specificity of *A. xylosoxidans* with respect to the distribution of clonal types. However, larger scale studies with samples from different clinical, environmental and geographical sources are needed to better understand and characterise the global population of this pathogen. This chapter demonstrated the determination of methods for identification of *A. xylosoxidans*. Molecular-based approaches, including 16S rDNA and MLST, showed relationships between isolates. However, this did not explain entire genetic relationship between these isolates. This emphasised that the construction of whole genomic relationship was required.

3.5. Conclusion and future work

The main purposes of this study were to demonstrate the ability of multiple bacterial identification approaches to identify emerging pathogens and to apply strain typing of clinical isolates of *A. xylosoxidans* in order to investigate geographical associations. Due to the fact that *A. xylosoxidans* is one of the most important emerging pathogens that cause severe diseases in Cystic Fibrosis patients and other immune-compromised patients, precise identification is required for epidemiological study and clinical work-up in order to deliver an appropriate treatment to the patients.

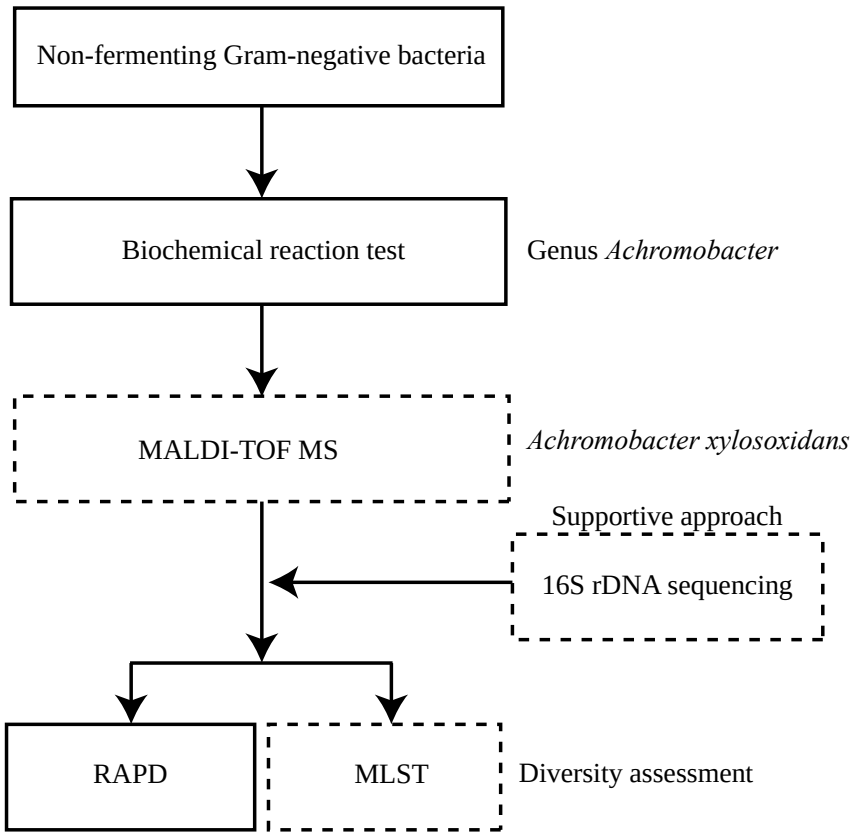


Figure 3.7: The proposed algorithm for *A. xylosoxidans* identification from non-fermenting Gram-negative. Dashed boxes indicate approaches that are affected by the lack of advanced facilities such as MALDI-TOF MS and sequencing facility.

With respect to reference labs, the identifications in the reference labs are merely based on biochemical (conventional) tests. MALDI-TOF MS is used in particular samples, such as samples that cannot be clearly identified by means of conventional methods. A number of identification methods were compared in this study. On the basis of the results presented here, a pragmatic algorithm for *A. xylosoxidans* identification is suggested (Figure 3.7).

Starting with non-fermenting Gram-negative bacteria from MacConkey agar, biochemical reaction-based methods, such as conventional tests and API20 NE can be used as screening tests due to their good performance in genus identification. For species identification, MALDI-TOF MS can be a main method to assign species to the isolates. The gene sequencing of 16S rDNA can be a supportive information for species identification. Finally, the intra-species discrimination can be conducted by additional methods, including MLST analysis and RAPD, which offer the potential to analyse within-species diversity.

Theorically, utilising a single approach that can give a highly reliable result is the best practice in clinical and diagnostic microbiology but the ideal approach is still unavailable. Therefore, using multiple approaches is necessary for current situation because results from several methods can reciprocally support one another to define a final result. However, some identification approachs, such as MALDI-TOF and DNA sequencing, cannot be used in a routine identification practice because some developing countries, such as Thailand, cannot afford those facilities for a routine purpose (They can afford for research purpose). Therefore, phenotypic identification remains a routine diagnostic practice in developing countries as the method still perform well in the identification of common pathogens, such as *S. aereus*, *E. coli*, and *S. pneumoniae.* Although the time-to-result of the phenotypic is approximately, at least, 48 hours, compared to its of MALDI-TOF which takes around 36 hours, the treatment is not affected because the treatment for infections, mostly, relies on antibiotic susceptibility of the pathogens.

However, there is still in need for a better and more robust identification method for the healthcare unit where MALDI-TOF MS and sequencing facility are

unavailable because rapid and accurate diagnostic provides the opportunity to establish empirical treatment, especially for particular pathogens having intrinsic resistance. Likewise, correct species identification is important for microbiological research and epidemiology. Whole genome sequencing of all *Achromobacter* isolates will be conducted to achieve species identification and to pursue futher study.

Chapter 4

Pan-genome analysis of *Achromobacter xylosoxidans*

4.1. Introduction

The previous analysis in Chapter 3 suggests that phylogenetic trees built using the 16S rDNA gene (Figure 3.4) and MLST genes (Figure 3.7) allow for *Achromobacter* species identification. However, sub-species level identification has not been clearly elucidated by those analyses. Further analysis of whole genome sequence data would enable a higher resolution of core phylogenomic analysis to be carried out, define the genome structure and pan-genome of *A. xylosoxidans*, and allow us to elucidate the molecular relationships between isolates.

4.1.1. Comparative genomic analysis in pathogenic bacteria

The use of comparative genomics to study bacterial evolution has revealed how bacteria undergo adaptation to survive in the host environment and how members of certain species from various niches can differ, particularly in genetic contents (Bentley & Parkhill, 2004). The comparative analysis of bacterial species is becoming important to reveal how a pathogen survives or causes diseases. For example, the pressure of host immune response pulses the adaptation of *E. coli* to be macrophage resistant strain, which is virulent to hosts (Miskinyte *et al.*, 2013). The mutation of *cis*-regulatory element allows host-dependent pathogenic adaptation of *Salmonella* (Osborne *et al.*, 2009).

Since the first genome of a pathogenic bacterium, *H. influenzae* Rd, was published (Fleischmann *et al.*, 1995), studies have used comparative genomic approaches to elucidate molecular and genetic factors that cause diseases in hosts.

One of the applications of comparative genomic analyses is to investigate the gain/loss of genes or genetic adaptations within the genomes of pathogens. An example that shows genetic adaptations leading to more virulent bacteria is seen in the discovery of extra plasmids in the *Bacillus cereus* group. Whilst the analysis showed that the genomes of *B. cereus*, an opportunistic pathogen that usually causes carbohydrate-related food-poisoning (Gilbert, Stringer & Peace, 1974; Schoeni & Lee Wong, 2005), *B. thuringiensis*, a biological insecticide (Bravo *et al.*, 2011), and *B. anthracis*, an anthrax-causing bacterium (Spencer, 2003), are similar, some genes encoding invasion factors and the capsule were, surprisingly, not found in the *B. anthracis* and *B. thuringiensis* genomes. Subsequent whole genome sequencing enabled the analysis of the complete genomes of *B. cereus*, *B. anthracis* and *B. thuringiensis* revealing 85% similarity between these genomes. As a result of the analysis, species-specific genes were identified to allow differentiation between *B. cereus* and *B. anthracis* on a molecular level (Ivanova *et al.*, 2003). Another example is the comparative genomic study of three pathogenic bacterial species in the genus *Bordetella* - *B. pertussis*, *B. parapertussis* and *B. bronchiseptica*. Adaptations in these species, in particular *B. pertussis* and *B. parapertussis*, occurred independently from each other and from their *B. bronchiseptica*-like ancestral, and the differences between the genomes of *B. pertussis* and *B. parapertussis* are regulated by the loss and the inactivation of the genes that results in different host-pathogen interaction (Parkhill *et al.*, 2003).

Comparative genomic analyses have also been used to demonstrate the adaptations that allow bacteria to survive in different environments. For example, the genomic study of *Stenotrophomonas* species illustrated host-specific genes (Alavi *et al.*, 2014). Genes associated with plant growth promotion were found only in plant-associated *S. rhizophila*, not in *S. maltophila*, whereas genes associated with virulence in humans were found only in the pathogenic *S. maltophila*. However, there is no significant genomic difference between plant-associated *S. maltophila* and pathogenic *S. maltophila*. In addition to determination of the genomes in different species, comparative genomics can shed the light on genomic difference between strains of the same species. For instance, lateral genetic transfers can mediate genetic variations in *A. baumannii*,

leading to antibiotic resistant phenotype of the species (Liu *et al.*, 2014). Comparative analysis can also reveal genetic variation of the pathogen from similar area, as illustrated by genetic variation of *Burkholderia pseudomallei*, a pathogen that causes melioidosis, in a rice paddy in northeast Thailand (Wuthiekanun *et al.*, 2009). Moreover, the significant change in genetic components of closely related species can be explored by comparative genome. For example, the pseudogenisation of gastroenteritis-associated genes in *S. typhi* serovars allows *S. typhi* to become human-specific strains and cause only systemic diseases, compared to *S. enterica* that have a broad host range and cause both gastroenteritis and systemic diseases (McClelland *et al.*, 2004). Also, another example is the genomic reduction in the genome of *B. mallei*, compared to closely-related species *B. pseudomallei*, led to the lack of genetic diversity of *B. mallei* (Losada *et al.*, 2010). All in all, these examples demonstrate the usefulness of comparative genomic analyses in understanding the biology of an organism.

4.1.2. Genetic element transfer and recombination

DNA can transfer between strains or species by three mechanisms: transformation, conjugation, and transduction (Frost *et al.*, 2005). Transformation was the transfer of bacterial DNA between closely related species by means of taking free DNA from the environment (Thomas & Nielsen, 2005). Inversely, another two mechanism require special tools to mediate DNA transfer. Conjugative transfer requires pairing organ called ‘Pilus’ that provides a route to transfer DNA (plasmid or integrative conjugative element) from a ‘donor’ cell to a ‘recipient’ cell (Ou & Anderson, 1970; Burrus & Waldor, 2004). The last possible mechanism of genetic transfer is conduction. As a mediator to assist genetic transfer, a bacteriophage, a bacterial virus, uses a benefit from host’s DNA replication machinery to replicate itself. As such, bacteriophages carry host’s DNA and transfer it to other bacteria where bacteriophages infect (Canchaya *et al.*, 2003). Once this has happened, DNA recombination events can occur.

Recombination can happen either at both the inter-species and intra-species level. Inter-species recombination is the exchange of genetic element that happens between species; for example, the recombination of glutamine synthetase and shikimate dehydrogenase genes between *Neisseria meningitidis* and other *Neisseria* species affects phylogenetic structure (Zhou, Bowler & Spratt, 1997). The effect of intra-species recombination varies between species but, mostly, their phylogenies are affected by the recombination (Feil *et al.*, 2001). Several studies have shown that recombination has been involved in outbreaks in certain species. For example, genetic recombination significantly accelerates evolution and adaptation in *Legionella pneumophila*, and the recombination contributes to 11-year outbreak in spain (Sánchez-Busó *et al.*, 2014). Recombination also has the same effect on host adaptation in *E. faecium* to survive in hospital environment with the acquisition of cell wall biosynthesis genes and drug resistance genes allowing it to survive in a hospital environment (de Been *et al.*, 2013). In the case of antibiotic resistance and epidemiological problems, recombination has a vast impact, in particular, on the spread of drug-resistant bacteria in the community. A recent study has shown that the spread of *S. pneumoniae* in a refugee camp was driven by genetic recombination (Chewapreecha *et al.*, 2014). In general, research into pneumococcal infections focuses on capsulated *S. pneumoniae* due to the virulence of the capsule (AlonsoDeVelasco *et al.*, 1995; Kadioglu *et al.*, 2008). Nevertheless, non-capsulated strains of *S. pneumoniae* that are neglected by many researchers, contribute to the transfer of antibiotic resistance gene, as determined by high rate of exchange of genetic elements (Chewapreecha *et al.*, 2014). Moreover, *S. pneumoniae* can obtain genetic element via transformation (Piotrowski, Luo & Morrison, 2009).

4.1.3. Microbial pan-genomes

An extension of a comparative genomic study is to investigate a comprehensive characteristic set of genomes, or ‘pan-genome’, in a species. The study of pan-genomes has elucidated the structure of bacterial species’ genome beyond what is achievable by studying individual genome sequences. There is no absolute

definition of the term ‘pan-genome’; however, a pioneering study by Tettelin *et al.* (2005) explained that the study of a pan-genome includes the determination of a core gene set and a disposable gene set with an explanation and extrapolation provided by a mathematical model. In the analysis of pan-genomes, a core genome is defined as ‘gene sets shared by all strains in the same species’, and a dispensable/accessory/disposable genome includes ‘gene sets present in 2 or more strains, but not all, in the same species’ and ‘gene sets specific to individual strains’ (Tettelin *et al.*, 2008).

In practice, the estimation of pan-genome is based upon the level of protein sequence similarity or gene similarity that can be determined by using various methods, such as gene alignment (Tettelin *et al.*, 2005), all-against-all BLAST searches (Jacobsen *et al.*, 2011), Cluster of Orthologous Groups (COGs) databases (Conlan *et al.*, 2012) and homology analyses (Soares *et al.*, 2013). The number of genes added was taken from responsible added genomes. The extrapolation of the size of core genome and the size of strain-specific genomes follows an exponential decay (Tettelin *et al.*, 2008):

$$N = \kappa e^{-n/\tau} + \Omega$$

Where N is the size of the pan-genome, n is the number of genomes compared, and κ, τ and Ω are free parameters.

Heaps’ law has been applied as a model for pan-genome estimation. In information retrieval, the principle of Heaps’ law is used to count the number of unique characters/words or ‘attributes’ in each document, or ‘entity’, collected. According to Heaps’ law, when more documents are collected, there will be some words that are repeated and some new words that are distinct (Egghe, 2007). The relationship between the number of words/characters, *T*, and documents collected, *A*, is that T varies as A^{γ}; where γ is a constant and $\gamma < 1$. Likewise, in genome science, as more genomes are retrieved, it is more difficult to find new genes (Tettelin *et al.*, 2008).

Heaps' law is applied to estimate the number of genes that will be found when a certain number of genomes are added (Tettelin *et al.*, 2008), as follows:

$$N = \kappa n^{\gamma}$$

Where N is the size of the pan-genome, n is the number of genomes compared, and κ and γ are free parameters.

In addition to the pan-genome's size, one can address the fact that the number of new attributes and the number of entities added also follow Heaps' law (Tettelin *et al.*, 2008). Therefore, the extrapolation of newly discovered genes can be used to explain the pan-genome with the following relationship:

$$N = An^{-\alpha}$$

Where N is the number of new genes discovered, n is the number of genome added, and A and α are free parameters.

Considering the relationship between the number of new genes discovered and the number of genomes added, when $\alpha > 1$, the number of new genes decreases dramatically to reach zero. This means that the size of the pan-genome nearly approaches a stable number, resulting in it being a 'closed-genome'. On the other hand, when $\alpha \leq 1$, the number of new genes is still detectable. In other words, the pan-genome is still increasing in size and is an 'opened pan-genome' (Tettelin *et al.*, 2008).

So far, pan-genome analysis has been performed in a number of pathogenic bacteria, for example, *Streptococcus agalactiae* (Tettelin *et al.*, 2005), *Haemophilus influenzae* (Hogg *et al.*, 2007), *Neisseria meningitidis* (Schoen *et al.*, 2008), *Escherichia coli* (Rasko *et al.*, 2008), *Yersinia pestis* (Eppinger *et al.*, 2010), *Streptococcus pneumoniae* (Donati *et al.*, 2010), *Bifidobacterium spp.* (Bottacini *et al.*, 2010), *Elizabethkingia anophelis* (Teo *et al.*, 2014), *Bacillus anthracis* (Rouli *et al.*, 2014), and *Haemophilus parasuis* (Howell *et al.*, 2014).

An example of a closed-genome organism is *B. anthracis*, which its core genome of which contributes to 99% of the whole genome, which suggests that *B. anthracis* is a pathogenic bacterium that does not need to retrieve external virulence genes (Rouli *et al.*, 2014). *S. pneumoniae* is a clear example of an open-genome species. Its open pan-genome explains how the species can adapt and respond to the environment rapidly (Donati *et al.*, 2010). It is a straightforward explanation that the adaptation of *S. pneumoniae* from non-virulence to virulence was driven by the acquisition of capsule genes.

4.1.4. Orthology and Paralogy of the genes

Thousands of bacterial genomes have been sequenced; however, relatively few genes have been well studied and characterised to find their function. Homology, by definition, describes the relationship between genes that share a common origin. Genes with a homologous relationship are called homologous genes or homologues (Koonin, 2005). Orthology, meanwhile, defines the homologous relationship that emerges from a speciation event of compared species. Genes that have orthologous relationships are called orthologous genes or orthologues. Another extension of homology is paralogy. Paralogy is a form of homologous relationship that happens via duplication within a species. Genes that have paralogous relationships are called paralogous genes or paralogues. A classical example of paralogous genes can be seen in the presence of adult haemoglobin genes and foetal haemoglobin genes in humans (Hardison, 2012). Whereas orthologous genes have equivalent functions, paralogous genes often have different functions despite the fact that their mechanisms of action are similar. For instance, the oxygen entrapment of foetal haemoglobin is more rapid than the oxygen entrapment of adult haemoglobin (Berg, Tymoczko & Stryer, 2002).

The use of orthologues and paralogues has been increasing and is becoming an important approach in genomics, especially in comparative genomic studies. A noticeable example of the application of orthologues/paralogues is a database called COGs (Tatusov, 1997) where large protein databases are collected and categorised into 25 sub-categories under four main categories. Recently, a

database called ‘evolutionary genealogy of genes: Non-supervised Orthologous Groups’ (eggNOG) has been introduced as a comprehensive database for orthologous group identification (Jensen *et al.*, 2008; Powell *et al.*, 2014). Currently, the eggNOG database version 4.0 includes more than 11 million protein covering 3686 organisms. The eggNOG also applies COGs categories to the annotation.

Besides database construction, orthologues/paralogues are also used to generate a core body in order to investigate shared genes and accessory genes amongst genomes analysed. For instance, they have been used to identify shared genes between *Bifidobacterium spp.* and other gut bacteria (Bottacini *et al.*, 2010), a core genome for recombination analysis in *E. faecium* (de Been *et al.*, 2013), a guide for gene annotation (Richardson & Watson, 2013) and a scaffold for whole genome phylogeny construction (Bertels *et al.*, 2014).

4.1.5. Comparative genomic analysis of the genus *Achromobacter*

A recent comparative study of six genomes within the genus *Achromobacter* genomes revealed that *Achromobacter* has a large open-genome and a low recombination rate (Li *et al.*, 2013). The genome size of members of genus *Achromobacter* is generally 6-7 Mbp with a GC content of 65-68%. Pan-genome analysis indicated that the size of the pan-genome is twice the average size of the genomes. This suggests the ability to be a reservoir and a donor for horizontal gene transfer. The size of the core genome is approximately half of the average genome size in the genus. Moreover, analysis of the core genome detected a low level of recombination in the core genome regions of *Achromobacter*.

4.1.6. Data mining using random forest, an unsupervised learning machine

The study of genome science is encouraged because of the development of genome sequencing technology and computational biology. However, the transformation of enlightened genome information into living organisms

phenotype is crucial. The establishment of the correlation between genome and biology is a challenge to genome scientists and biologists. To deal with large-scale genomic data and biological expression, untrained or unsupervised machine learning principle has been applied to link genomic information to biology. Random Forest™ is one of the famous unsupervised learning algorithms that are used to operate data mining (Breiman, 2001). The theory behind Random Forest™ is that the machine constructs a binary tree using random input variables to from each daughter node. Then the algorithm groups the data based on the tree and classification indicated. The tree is re-drawn 'n' times for bootstrap support purpose. The measurement of variable importance is a special feature of Random Forest™. Given an example of analysis in genotype-phenotype association, input variables can be the presence of SNPs or genes in the genomes, and the classification can be phenotypes of interest. The machine randomly selects genes or SNPs to construct a binary tree, and, subsequently, group genomes with respect to the existence of variables and phenotypes. Finally, the variables are ranked on the basis of the purity of classification. In this study, Random forest™ was applied as a main data mining approach to correlate genomic information with phenotype information.

4.1.7. Aims and objectives

The complete genome sequence of *A. xylosoxidans* NH44784-1996 elucidated an understanding of the molecular and genetic content of *A. xylosoxidans* (Jakobsen *et al.*, 2013). In addition, a subsequent analysis revealed better knowledge of potential antibiotic resistance genes in the species (Hu *et al.*, 2015). However, a comprehensive analysis of the species, especially from diverse geographical sources, has not been carried out. Therefore, the first aim of this study is to investigate whether *A. xylosoxidans* strains from different continents are genetically different. In order to reveal the population structure of *A. xylosoxidans* isolates, next-generation sequencing technologies were utilised to sequence the genomes of the isolates and compare these new sequences to the recently published reference genome, *A. xylosoxidans* NH44784-1996. Gene orthology and phylogenetic relationships will be used as the main approaches to

compare these genome sequences to one another. Furthermore, a recombination analysis will be conducted to investigate whether recombination events affect the phylogenetic structure of *A. xylosoxidans*. By means of the comparative genomic analysis, the understanding of *A. xylosoxidans* in terms of the genetic relationships between its strains will be more enlightened.

As well as its population structure, many genomic characteristics of *A. xylosoxidans* also remain unclear. According to a study by Hu *et al.* (2015), *A. xylosoxidans* genomes are likely to be receptive genomes because of the existence of mobile genetic elements in *A. xylosoxidans* genome. This indicates that the pathogens can exchange genetic elements with their surroundings. To unveil and investigate the genome structure of the species, an analysis of its pan-genome will be pursued using newly sequenced genomes together with a reference genome. By doing so, an estimation of genes encoded by the entire *A. xylosoxidans* species, as well as an estimation of numbers of common genes, will be achievable to tell us whether genomes of *A. xylosoxidans* behave more like they are closed or open.

4.2. Materials and methods

4.2.1. Genome mapping of sequence reads and genome assembly

The first step in a comparative genomic analysis is to assess the similarity between sequence genomes and the reference genome. Here, the genomes of the 38 strains, consisting of 31 clinical strains and seven reference strains, were sequenced using short-read sequencing on the Illumina HiSeq 2000. Adapters were trimmed from the sequenced reads using Cutadapt version 1.1 (Martin, 2011) with an option – O 3 - that trimmed 3' ends of the reads that matched the adapters. Then, reads with base quality values below 20 were trimmed using Sickle version 1.200 (Joshi & Fass, 2011). Mapping of processed reads to the reference genome, *A. xylosoxidans* NH44784-1996, was performed using the Burrows-Wheeler Aligner (BWA) software version 0.5.9-r16 (Li & Durbin, 2009), which uses the Burrows-Wheeler transformation to identify matches between reads and the reference sequence. Duplicated reads were removed using Picard version 1.85 (http://picard.sourceforge.net/), using the following command format:

java -jar MarkDuplicates.jar ASSUME_SORTED=true MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=1000 I=<original mapping file in BAM format> O=<duplication removed file in BAM format> REMOVE_DUPLICATES=true VALIDATION_STRINGENCY=LENIENT

The mapping percentage of sequenced strains to the reference was obtained using the 'flagstat' option in SAMtools version 0.1.18-r580 (Li & Durbin, 2009), enabled using the following command:

samtools flagstat <alignment file in BAM format>

Polymorphisms obtained from read mapping were extracted and were used to generate pseudoalignment using Snippy version 2.5 (http://www.vicbioinformatics.com/software.snippy.shtml). A phylogenetic tree

was then constructed using PhyML version 3.1 to visualise a relationship of the strains using *P. aeruginosa* PAO1 as an outgroup.

Prior to *de novo* assembly, k-mer values of sequence reads from each strain were estimated to attend optimum N50 values, using VelvetOptimizer version 2.2.5 (Gladman & Seemann, 2008). The following command, testing k-mer values between19 and 89, is an example of commands used:

Perl VelvetOptimiser.pl -s 19 -e 89 -f '-fastq –shortPaired <1st read file> -shortPaired2 <2nd read file>'

These k-mer values obtained from this optimisation step were used in *de novo* assemblies to generate draft genomes of the strains. Genome assemblies, using the De Bruijin graph algorithm, were performed with Velvet version 1.2.07 (Zerbino & Birney, 2008) using the following command format:

velveth <output_file> <k-mer> -fastq -shortPaired <1st read file> -shortPaired2 <2nd read file>

velvetg

In addition to the Velvet assembler, *de novo* genome assemblies were executed using Spades version 3.1.1 (Bankevich *et al.*, 2012). Spades is another short read assembler that uses the De Bruijin graph algorithm but its advantage is that it can handle multiple k-mer values at the same time. *De novo* assemblies using Spades were computed using the following command format:

spades.py -k <list of k-mer separated by comma > --careful --pe1-1 <1st read file> --pe1-2 <2nd read file> -o <output file>

The list of k-mer values used in the above command was '51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81'.

4.2.2. Gene annotation and functional assignment

The next step in a comparative genomic analysis is to predict gene models within the draft genomes. Rapid annotation software, Prokka version 1.7.2, was used to predict and annotate genes within the draft genomes in this study (Seemann, 2014). Firstly, Prokka predicted coding sequences in the draft genomes using 'Prodigal' (Hyatt *et al.*, 2010). Prodigal's algorithm looks for all open reading frames then creates hexamers for each of these sequences. The score given to each gene is calculated from the hexamers' coding scores. After creating potential coding sequences, the software checks for ribosome-binding sites based on Shine-Dalgano sequences, as a means of analysing the likelihood that identified sequenced could be translated. Then, connections between nodes/genes that are calculated before a final set of results, comprised of amino acid sequences of encoded by each gene, are generated. In parallel, tRNA genes and rRNA operons were predicted using tRNAscan-SE (Schattner, Brooks & Lowe, 2005) and RNAmmer (Lagesen *et al.*, 2007), respectively. After the gene calling process, the genes were annotated with a customised database, namely 'Gbbct', from which most updated bacterial genes from Genbank were obtained. Prokka was executed using the following command format:

prokka –locustag <strain's name> --outdir <output file> --prefix <strain's name> --addgenes –-usegenus –genus Gbbct <contig file>

The assignment of functional groups to genes was conducted by reciprocal protein BLAST comparison (Altschul *et al.*, 1990) to a bacterial gene pool using the eggNOG database version 4.0 (Powell *et al.*, 2014). The eggNOG classifies genes based upon Clusters of Orthologous Groups (COGs) (Tatusov, 1997). A command used to perform reciprocal BLAST searches against the eggNOG database is shown in Appendix 1.1. Functional annotation by eggNOG follows COGs categories categorised into 25 sub-categories under four main categories, as follows:

Table 4.1: Functional annotation defined by COGs.

Category		Abbreviation
Information storage and processing	RNA processing and modification	(A)
	Chromatin structure and dynamics	(B)
	Translation, ribosomal structure and biogenesis	(J)
	Transcription	(K)
	Replication, recombination and repair	(L)
Cellular processes	Cell cycle control, cell division and chromosome partitioning	(D)
	Cell wall/membrane biogenesis	(M)
	Cell motility	(N)
	Posttranslational modification, protein turnover and chaperones	(O)
	Signal transduction mechanisms	(T)
	Intracellular trafficking, secretion and vesicular transport	(U)
	Defense mechanisms	(V)
	Extracellular structures	(W)
	Cytoskeleton	(Z)
Metabolism	Energy production and conversion	(C)
	Amino acid transport and metabolism	(E)
	Nucleotide transport and metabolism	(F)
	Carbohydrate transport and metabolism	(G)
	Coenzyme transport and metabolism	(H)
	Lipid transport and metabolism	(I)
	Inorganic ion transport and metabolism	(P)
	Secondary metabolites transport and metabolism	(Q)
Poorly characterised	General function prediction	(R)
	Function unknown	(S)

4.2.3. Comparative genomic analysis

Of the '*A. xylosoxidans*' strains isolated for this project, the results of MLST analysis (Figure 3.7) and genome mapping suggested that only 25 were *A. xylosoxidans* – NCIMB 11015, DSM 2402, L1, L5, L8, L10, L11, L14, L15, L16, L17, R1, R2, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R13, R14. As such, only 26 genomes, including NH44784-1996, of *A. xylosoxidans* strains were used in the comparative genomic analysis.

The annotated proteins from these 26 *A. xylosoxidans* genomes were clustered into groups using an all-versus-all BLAST search with the Markov Cluster Algorithm (MCL) implemented in OrthoMCL version 1.4 (Li, Stoeckert & Roos, 2003). The following command format was used:

perl orthomcl.pl --mode 1 --fa_files <list of strains' genes, separated by commas>

OrthoMCL's output, showing gene orthologous groups, was parsed into a matrix showing the presence/absence of each orthologous group in each genome using in-house scripts (Shown in Appendix 1.2). The occurrence of the gene families was shown in binary format (1 = presence; 0 = absence). With the table representing strains in columns and orthologous groups in rows, core orthologous groups were defined as the groups that were present in every genome, and accessory orthologous groups were defined as the groups that were present in some genomes but not all.

To investigate the strains as clusters, based on the occurrence of gene orthologues, the table showing strains and OrthoMCL's output was subjected to hierarchical clustering and heat mapping using the 'gplots' package implemented in the R statistical package, version 3.1.2 (Team, 2014). In the cluster analysis, the distance calculation method and clustering method were 'binary' and 'complete', respectively, as per the following R command lines:

dist<- as.matrix(dist(tab, method="binary"))
heatmap.2(dist.bin)

4.2.4. Pan-genome analysis

A mathematical model was applied to predict the sizes of the pan-genome and core genome through extrapolation. Randomly selected gene families were used to simulate how the number of genes increased and decreased when genomes were added. One thousand permutations were applied on the simulation to ensure

that there was no bias when genes were picked. The medians were used as central tendency to estimate least-squares fit.

The pan-genome analysis was performed by the addition of 26 genomes of *A. xylosoxidans*. The regression analysis for the pan genome and new genes discovered was executed by fitting a power model (Heaps' law) to the median of the data (Tettelin *et al.*, 2008):

$$N = \kappa n^{\gamma}$$

Where N is the size of the pan genome, n is the number of genomes, and κ and γ are free parameters.

The core genome analysis was performed by adding 26 genomes. The regression analysis for the core genome was executed by fitting an exponential model to the median of the data:

$$N = \kappa e^{-n/\tau} + \Omega$$

Where N is the size of the pan-genome, n is the number of genome, and κ, τ and Ω are free parameters (Conlan *et al.*, 2012).

Additional analysis of newly discovered genes was conducted in order to support an evidence of a pan-genome. The analysis was performed with the addition of *A. xylosoxidans* genomes and the number of new genes with each addition was estimated. The median of the number of new genes was fitted to a power model as per following formula:

$$N = An^{-\alpha}$$

Where N is the number of new genes discovered, n is the number of genome added, and A and α are free parameters.

Pan-genome and core genome extrapolation were both conducted in R with scripts and command lines courtesy of Dr. Jennifer Kelly (Appendix 1.3).

4.2.5. Phylogenetic tree construction

Phylogenetic relationships between the *A. xylosoxidans* strains were estimated using the core genome. In this study, the core genome was defined as the gene sets in the OrthoMCL output that contained exactly one gene from each strain. The corresponding nucleotide sequences were extracted using in-house Perl scripts and were aligned using MUSCLE (Edgar, 2004) with the following command type:

muscle -in <input file> -out <output file>

All gaps in the alignment were then stripped using trimAl software (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009) with and option - '–gt 1' - that removes all columns with gaps from the alignments. The format of the Trimal command line used is follows:

Trimal –in <input alignment file(gaps intact)> -out <output file(gaps removed)> -gt 1

After alignment concatenation, a phylogenetic tree was constructed using Mr. Bayes version 3.2.3, which uses Bayesian inference tree construction with a Markov Chain Monte Carlo (MCMC) algorithm to estimate posterior probability (Ronquist & Huelsenbeck, 2003). A generalised time reversible substitution model with 100,000 generations and ten sampling frequencies was used in the MCMC analysis. A phylogenetic tree was also constructed using a maximum-likelihood model, with 1,000 bootstrap values, in PhyML version 3.1 (Guindon *et al.*, 2010). The phylogenetic trees were viewed in FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree).

4.2.6. Recombination detection

Recombination can accelerate bacterial evolution; however, recombined sequences can interfere with bacterial phylogeny. To identify recombination signals in the core genome of *A. xylosoxidans*, multiple softwares, including a Neighbor-Net phylogeny investigation, Pairwise Homology Index test (PHI) tests and Bayesian inference-based analyses in ClonalFrame (Didelot & Falush, 2007), BratNextGen (Marttinen *et al.*, 2012) and ClonalFrameML (Didelot & Wilson, 2015), were applied.

The estimation of core genome alignment began with the Neighbor-Net phylogenetic tree construction and PHI test. These processes were performed using SplitTree (Bryant & Moulton, 2004). Following the detection of a significant recombination event using the PHI test, a recombination analysis was executed with ClonalFrame first. The software uses a MCMC algorithm together with Bayesian inference model to identify homologous recombination. Unfortunately, the analysis ran for nearly a month without finishing, so the analysis was stopped and performed using another software. BratNextGen was used to detect recombination in the core genome alignment. The principle of BratNextGen is to apply a Bayesian inference model to investigate homologous recombination in the genome alignment.

Although BratNextGen and ClonalFrame both use Bayesian inference models, BratNextGen is believed to be able to handle large datasets, such as whole genomes, better than ClonalFrame, which is more suitable for smaller datasets, such as MLST analyses. Initially, all default parameters were used in the analysis but this did not allow for clustering in BratNextGen's output tree. We sent a query to Dr. Pekka Martinen, an author and co-creator of the BratNextGen software. He advised setting the alpha-parameter to '1' when working with a diverse collection of strains. Hence, the analysis subsequently used default settings apart from the alpha value, which was set to '1'. The test of the significance of recombination signal-containing regions was executed with 100 permutations on a computer cluster. A threshold of 5% was set as a significant level for each suspicious region. Furthermore, ClonalFrameML, new recombination analysis software that uses maximum-likelihood algorithm, was applied to search for recombination in the core genome of *A. xylosoxidans*.

4.2.7. Variable classification using Random Forest™

A matrix of gene occurrence of *A. xylosoxidans* isolates was used as an input for the analysis. Random Forest™ was conducted using 'randomForest' package, implemented in R. The best 'mtry' parameter was pre-optimised using a built-in command 'tuneRF'. The analysis was performed with following parameters: ntree = 5,000 and pre-optimised 'mtry'. The input was prepared as a matrix presenting the occurrence of the genes in each genome and phenotypes of each genome. The result of the analysis was visualised using a command 'varImpPlot', which orders variables by variable importance score and shows the variables in a plot format.

4.3. Results

4.3.1. Genome mapping

All sequence reads from 38 *Achromobacter* isolates were mapped against a published reference genome, *A. xylosoxidans* NH44784-1996, using BWA. Twelve strains that were found not to be *A. xylosoxidans* through MLST analysis (Figure 3.7), consisting of five non-*xylosoxidans* reference strains and seven clinical isolates from the U.K. In these twelve strains, less than 80% of reads successfully mapped to the reference genome. The mapping percentage of strain L6 was also less than 80% (Appendix Table A2.1).

Three *A. xylosoxidans* strains mapped to the genome of strain NH44784-1996 (Appendix Table A2.1), with mapping percentages higher than 90.0%. In 17 other strains, 85.0 to 89.9% of reads mapped to the reference strain. Sequenced reads from strain L1 reached the highest mapping percentage at 96.1%. The lowest mapping percentage was 80.8%, achieved for reads from L15. Two type strains, NCIMB 11015 and DSM 2402, mapped to the reference genome with 85.6% and 84.7% of their reads, respectively.

Reads from the Illumina sequencing platform were mapped to the newly sequenced genome of *A. xylosoxidans* strain NCIMB 11015, generated by the Pacific Bioscience platform (Appendix Table A2.1; Appendix 5). Self-mapping of strain NCIMB 11015 showed that 99.68% of the Illumina reads mapped, which is the highest of all of the strains tested. This self-mapping evaluated the quality of sequencing platform. The lowest percentage of a strain's reads to map to strain NCIMB 11015 was 76.1%, as seen in strain L15. Nine strains were mapped with percentages in between 85.0 and 89.9%. Thirteen strains fell in the range of 80.0 – 84.9%. In two strains, less than 80% of reads mapped to strain NCIMB 11015.

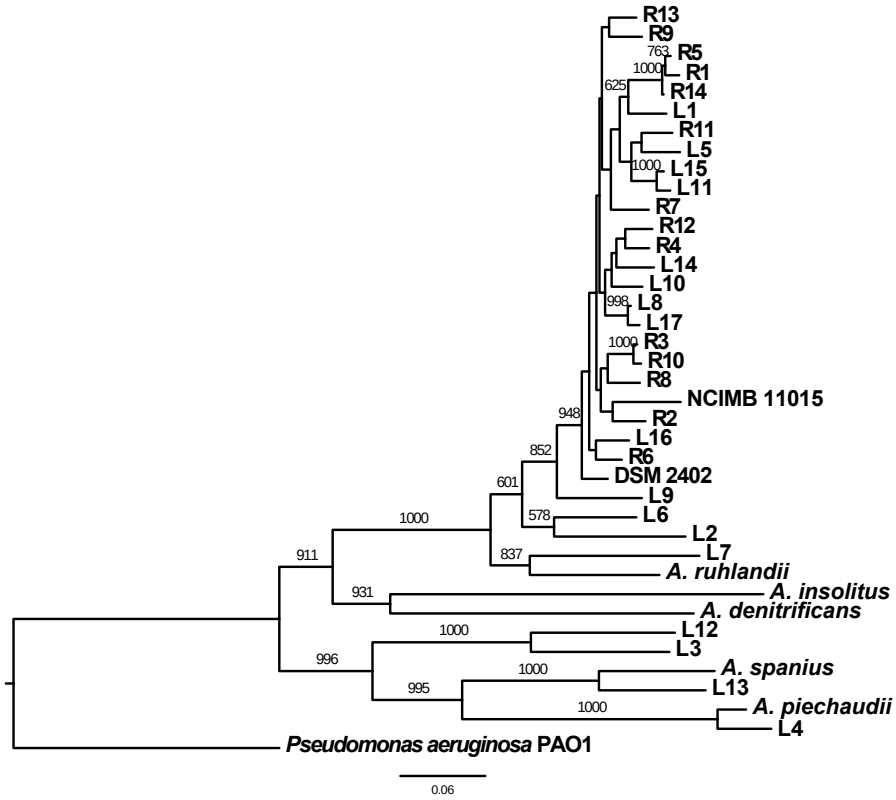


Figure 4.1: A maximum-likelihood phylogenetic tree constructed from 2,470 core genome SNPs using PhyML version 3.1. *P. aeruginosa* PAO1 is an outgroup. Strains L1-L17 are clinical isolates from the U.K. (mainly Liverpool). Strains R1-R14 are clinical isolates from Bangkok, Thailand. NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015 and DSM 2402 stands for *A. xylosoxidans* DSM 2402. Bootstrap values more than 50% were indicated.

Phylogenetic analysis of core SNPs presented a reliable cluster that contains *A. piechaudii*, *A. spanius*, L3, L4, L12, and L13 (>90% bootstrap supports). Some non-*xylosoxidans Achromobacter* species, *A. denitrificans*, *A. insolitus*, *A. ruhlandii*, and L7, were clustered closely to *A. xylosoxidans*. However, strains L2, L9, and, particularly, L6, which was ambiguously identified by multiple approaches (refer to Chapter 3), were clustered separately from the *A. xylosoxidans*'s cluster with moderate confidence (>60% bootstrap value) (Figure 4.1). Taken all together, this suggested that L6 was not considered to be an *A. xylosoxidans* strain. These 13 strains were excluded from this study. Therefore, only 25 isolates' genomes – 14 Thai isolates, 9 British isolates, and 2 reference strains – were used in subsequent analyses in this study. The nine British isolates

will be renamed ‘Liverpool’ isolates in this thesis, as all of them were from Liverpool.

Table 4.2: Short-read assembly statistics for all 25 *A. xylosoxidans* isolates used in this study, including pre-optimised k-mer values, numbers of contigs, sizes of largest contigs, N50 values, coverage values and the total sizes of contigs. Assemblies were performed using Velvet assembler version 1.2.07.

Strain	Assembly Information					
	K-mer	Contigs	Largest Contig (bp)	N50 (bp)	Coverage	Total (bp)
NCIMB 11015	61	720	53,541	8,063	> 100X	6,288,311
DSM 2402	65	695	94,834	29,946	> 100X	6,788,478
L1	69	418	180,149	53,488	> 100X	6,777,785
L5	59	2,245	88,090	13,937	> 100X	6,480,710
L8	73	904	150,975	22,877	> 100X	6,754,552
L10	69	1,061	124,477	24,316	> 100X	6,842,075
L11	75	806	144,777	24,225	> 100X	6,763,626
L14	63	1,852	117,346	16,809	> 100X	6,574,805
L15	67	2,381	61,892	7,605	> 100X	6,726,584
L16	73	917	105,215	22,497	> 100X	6,742,623
L17	67	891	153,800	32,174	> 100X	6,808,965
R1	63	1,106	96,079	25,454	> 100X	6,831,430
R2	65	1,148	117,644	21,480	> 100X	6,426,085
R3	57	2,650	45,043	5,996	> 100X	6,391,477
R4	61	1,627	83,204	10,449	> 100X	6,521,510
R5	57	1,468	138,145	13,767	> 100X	6,789,429
R6	61	1,946	84,664	9,768	> 100X	6,554,317
R7	65	2,380	49,203	6,457	> 100X	6,604,764
R8	63	2,006	43,169	8,471	> 100X	6,590,525
R9	63	2,537	48,754	6,275	> 100X	6,269,108
R10	61	2,319	40,301	6,943	> 100X	6,408,088
R11	65	961	117,063	17,165	> 100X	6,361,925
R12	65	1,044	106,951	18,417	> 100X	6,341,870
R13	67	1,285	148,572	12,642	> 100X	6,443,125
R14	67	1,153	114,801	14,264	> 100X	6,783,140
Mean	-	1,461	100,348	17,459	-	6,594,612
S.D.	-	672	40,240	11,026	-	190,068

Table 4.3: Short-read assembly statistics for all 25 *A. xylosoxidans* isolates used in this study, including numbers of contigs, sizes of largest contigs, N50 values, coverage values and the total sizes of contigs. Assemblies were performed using Spades assembler version 3.1.1.

Strain	Assembly Information				
	Contigs	Largest Contig (bp)	N50 (bp)	Coverage	Total (bp)
NCIMB 11015	860	141,469	17,424	> 100X	6,489,821
DSM 2402	127	685,532	130,241	> 100X	6,802,447
L1	77	1,037,584	629,784	> 100X	6,793,720
L5	191	508,865	92,693	> 100X	6,470,420
L8	110	625,007	196,476	> 100X	6,771,819
L10	151	714,992	170,117	> 100X	6,859,600
L11	114	581,453	178,548	> 100X	6,790,401
L14	156	335,678	126,442	> 100X	6,571,402
L15	287	435,904	60,726	> 100X	6,801,183
L16	142	525,536	130,061	> 100X	6,765,818
L17	102	753,594	241,184	> 100X	6,815,184
R1	184	445,795	130,265	> 100X	6,924,972
R2	133	462,894	135,643	> 100X	6,453,898
R3	1,132	107,851	11,612	> 100X	6,483,870
R4	544	199,593	29,940	> 100X	6,577,466
R5	480	222,108	32,361	> 100X	6,827,548
R6	520	241,447	27,501	> 100X	6,605,589
R7	908	153,959	16,343	> 100X	6,750,977
R8	593	153,097	25,188	> 100X	6,670,944
R9	603	103,277	25,145	> 100X	6,395,077
R10	783	143,979	18,672	> 100X	6,498,797
R11	281	257,991	52,469	> 100X	6,394,985
R12	255	233,326	64,934	> 100X	6,375,895
R13	328	397,578	48,338	> 100X	6,573,511
R14	346	236,334	47,456	> 100X	6,849,323
Mean	376	388,194	105,583	-	6,652,587
S.D.	296	244,132	127,413	-	172,543

4.3.2. Genome assembly and general characteristics of *A. xylosoxidans* genome

Short read assemblies of 25 *A. xylosoxidans* strains were performed using Velvet version 1.2.07, with pre-optimisation performed using VelvetOptimizer version 2.2.5 (Table 4.2) and Spades assembler version 3.1.1 (Table 4.3). With Velvet, the average assembled length of each genome sequenced was 6.6 Mb with a standard deviation of 0.2 Mb. The number of contigs varied from 418 – 2,650 contigs with an average of 1,461 contigs and a standard deviation of 672. The average N50 value was 17,459. The size of the largest contig in each assembly was, on average, 100,348 bp. Spades performed better than Velvet. With Spades the total size of the genomes assembled was, on average, 6.7 Mb with a standard deviation of 0.2 Mb. The mean of the number of contigs assembled was 376 contigs, which is less than was seen in of Velvet's assemblies. The average N50 value was higher, with 105,583 bp. The average size of the largest contigs was approximately four times that seen in Velvet's assemblies. The result suggested a better performance by the Spades assembler in terms of contigs' sizes and N50 values.

In terms of gene prediction, genome assemblies from the Velvet had an average GC content and standard deviation value of 67.5% and 0.16%, respectively (Table 4.4). The lowest GC content was 67.33%, which belonged to R5, and the highest GC content was 67.88%, which belonged to R8. An average of 6,094 coding sequences (CDS), with approximately 38 tRNA genes and 3 rRNA operons, was predicted in the draft genomes. In genome assemblies output by Spades assembler (Table 4.4), the average GC content was 67.54% and the standard deviation was 2%. The lowest and the highest GC content values were 62.04% and 69.88%, respectively. Prokka software was used to predict an average of 6,052 CDS with 63 tRNA genes and 3 rRNA operons.

Table 4.4: Genome features of all *A. xylosoxidans* isolates, including GC content, number of predicted CDS, number of predicted tRNA sequences and number of predicted rRNA sequences. These features were obtained from Prokka version 1.7.2.

Strain	Velvet assembler				Spades assembler			
	%GC	CDS	tRNA	rRNA	%GC	CDS	tRNA	rRNA
NCIMB 11015	67.49	5,871	33	3	69.23	5,915	65	3
DSM 2402	67.39	6,140	37	2	67.26	6,150	60	2
L1	67.47	6,251	31	3	62.04	6,225	54	3
L5	67.67	5,919	38	2	67.42	5,823	61	2
L8	67.55	6,139	38	3	64.44	6,111	61	3
L10	67.57	6,263	41	2	64.57	6,256	62	2
L11	67.37	6,277	41	3	65.56	6,232	59	3
L14	67.57	6,055	32	3	66.55	5,962	62	3
L15	67.33	6,414	33	3	68.19	6,264	61	3
L16	67.35	6,142	33	3	65.21	6,101	65	3
L17	67.45	6,150	45	3	64.97	6,153	59	3
R1	67.36	6,246	45	3	66.69	6,271	68	3
R2	67.80	5,836	40	3	67.01	5,778	66	3
R3	67.48	6,039	31	3	69.57	5,993	61	3
R4	67.65	6,041	29	3	69.10	5,988	65	3
R5	67.33	6,254	40	3	68.69	6,198	68	3
R6	67.55	6,099	42	3	69.12	6,041	69	3
R7	67.53	6,273	39	3	69.88	6,215	67	3
R8	67.52	6,206	27	3	69.01	6,124	64	3
R9	67.78	5,921	28	3	69.72	5,787	64	3
R10	67.66	6,003	32	3	69.88	5,935	63	3
R11	67.75	5,828	34	2	68.36	5,816	60	2
R12	67.88	5,816	40	3	68.09	5,771	65	3
R13	67.70	5,943	36	3	69.20	5,974	64	3
R14	67.37	6,215	38	3	68.70	6,210	68	3
Mean	67.54	6,094	38	3	67.54	6,052	63	3
S.D.	0.16	167.08	9	0	2	169.23	4	0
Minimum	67.33	5,816	27	2	62.04	5,771	54	2
Maximum	67.88	6,414	45	3	69.88	6,271	69	3
Range	-	598	18	1	-	500	15	1

Statistical analyses performed using Student's t-test showed a non-significant difference between the number of predicted CDS in Velvet assemblies and Spades assemblies (Student's t-test; P=0.383), but the number of tRNA genes predicted in Spades assemblies was significantly higher than in Velvet assemblies (Student's t-test; P=8.01E-25). The analysis suggested that Spades assembler performed better in assembling short reads by constructing larger and fewer contigs and a greater number of tRNA genes, which was comparable to the genome of strain NH44784-1996 (53 tRNAs). As a result, genome assemblies from the Spades assembler were used in subsequent analyses in this study. However, GC content of strain L1 from Spades assembler was the lowest (62.04%). The analysis of L1's reads which did not map to the genome of NH44784-1996 revealed the presence of the components of bacteriophage. This can cause low GC content in genome assembly of strain L1. Therefore, the genome of L1 would be used in further analysis with the caveat that the assembly had low GC content and the presence of phage's components.

4.3.3. Pan-genome analysis

To determine a broad picture of the genome, a pan-genome was applied and fitted to a mathematical model in R. The pan-genome of *A. xylosoxidans* was determined by sequentially adding the whole gene sets of each genome and fitting them to a power model following Heaps' law. At the beginning of the curve (Figure 4.2), the median number of orthologues from a randomly picked genome was approximately 6,000 orthologues. The median increased when genomes were added until it reached a maximum point. A fitted curve (Figure 4.2) predicted that the genome size of *A. xylosoxidans* reached 14,394 orthologous groups when 26 genomes were added. This was approximately 2.4 times the median of a single genome's size and 4 times the core genome's size. With the power law's (Heaps' law) estimation, the data fitted the power law with an exponent in the following model:

$N = 5571.59 \times n^{0.291315}$ ($F_{1,24} = 1802$; $R^2 = 0.99$; p-value < 2.2e-16)

Where N = the size of the pan-genome and n = the number of genomes.
With R^2 at 0.99, this model was able to explain 99% of the data, so the model could predict the size of the core genome.

By considering newly discovered orthologues, the number of unique orthologues found in each genome was 920 orthologous groups on average (Figure 4.3). With every single genome added, the number of new genes decreased and the total reached 242 genes when the 25th genome was added. With power law model fitted to medians, the curve of new genes followed a decreasing power law explained in the following equation:

$N = 838.6 \times n^{-0.38568}$ ($F_{1,23} = 1351$; $R^2 = 0.98$; p-value < 2.2e-16)

Where N = the number of unique orthologous genes added and n = the number of genomes added. According to the R^2-value, this model covered 98% of the data.

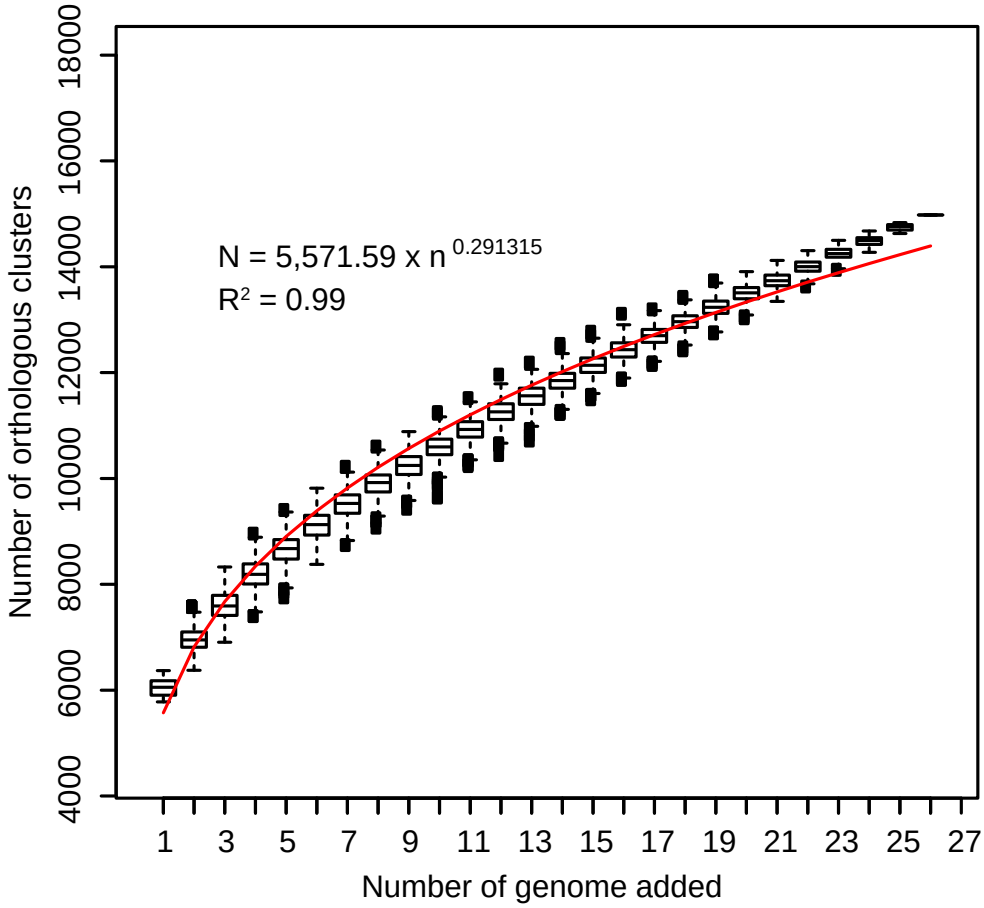


Figure 4.2: Boxplot showing the mathematical extrapolation curve of the *A. xylosoxidans* pan-genome. The analysis shows that the number of orthologues increases with the number of genomes added. The red line is a least squares fit of the power law to the median values. According to the power law, $N = \kappa n^{\gamma}$, so $\gamma > 0$ suggests an open-genome species.

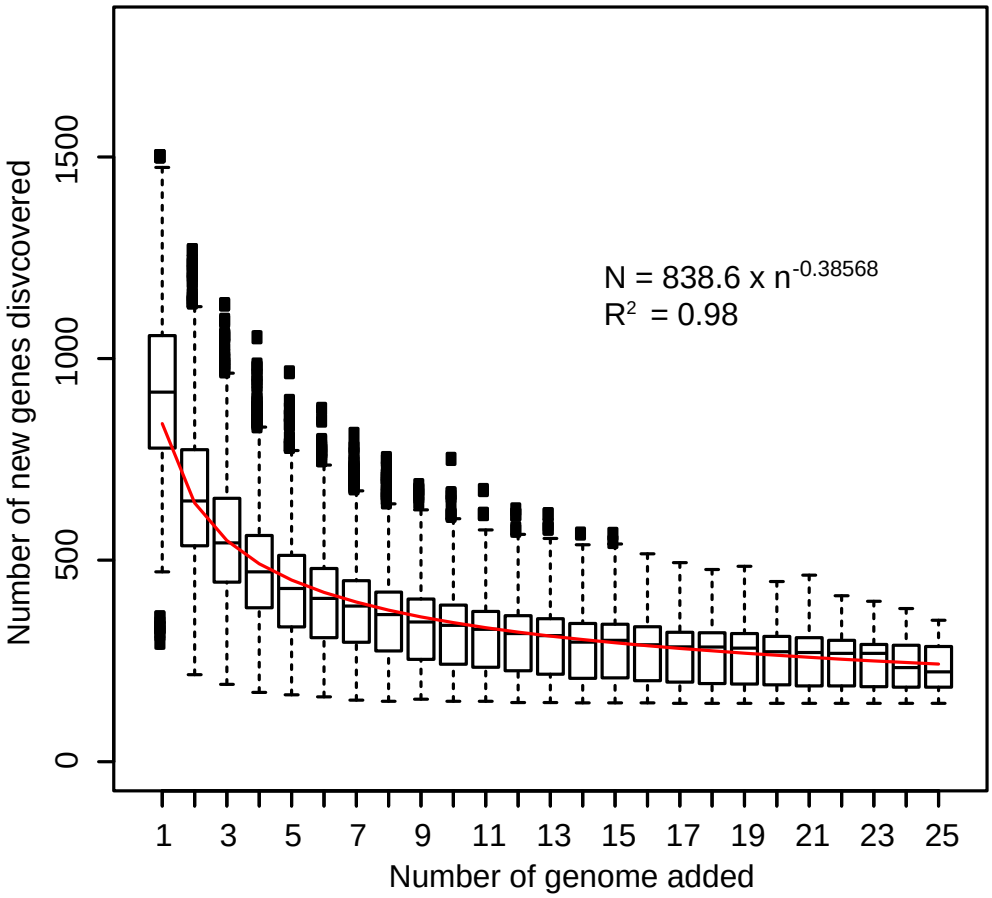


Figure 4.3: Boxplot showing the power model of new orthologues found with the additional of multiple *A. xylosoxidans* strains' genomes. The red line is a least squares fit of the power law to median values. According to the power law, $N = \kappa n^{-\alpha}$, so $\alpha \leq 1$ suggests an open-genome species.

4.3.4. Core genome analysis

The next questions to be answered concerned the size of the core genome and the functions to which the core genome contributed. As previously described, there were 3,764 commonly shared gene families recognised as the core orthologues of the *A. xylosoxidans* genome. The three largest orthologous groups consisted of 154 genes, 92 genes and 75 genes, respectively, from 26 genomes; they were RND-type efflux-coding genes, Acr-family proteins and dipeptide transport ATP binding proteins, respectively.

Using a mathematical approach, the result was, as expected, that the core genome's size decreased with each genome added. The median of the randomly selected first genome was approximately 6,000 orthologues and it sharply decreased to 5,000 when the second genome was added (Figure 4.4). The median number of gene orthologues then gradually reduced until it nearly reached an asymptote (a blue dashed line in Figure 4.4). The estimation of the exponential model with a least squares fit method showed that the curve reached a minimal value of 3,654 orthologous groups, which was approximately half of the genome in one genome, following this equation:

$$N = (1720.05 \times e^{-0.10030 \times n}) + 3{,}654 \ (F_{1,24}=970;\ R^2 = 0.98;\ \text{p-value} < 2.2e\text{-}16)$$

Where N = the number of orthologues in the core genome and n = the number of genomes added.

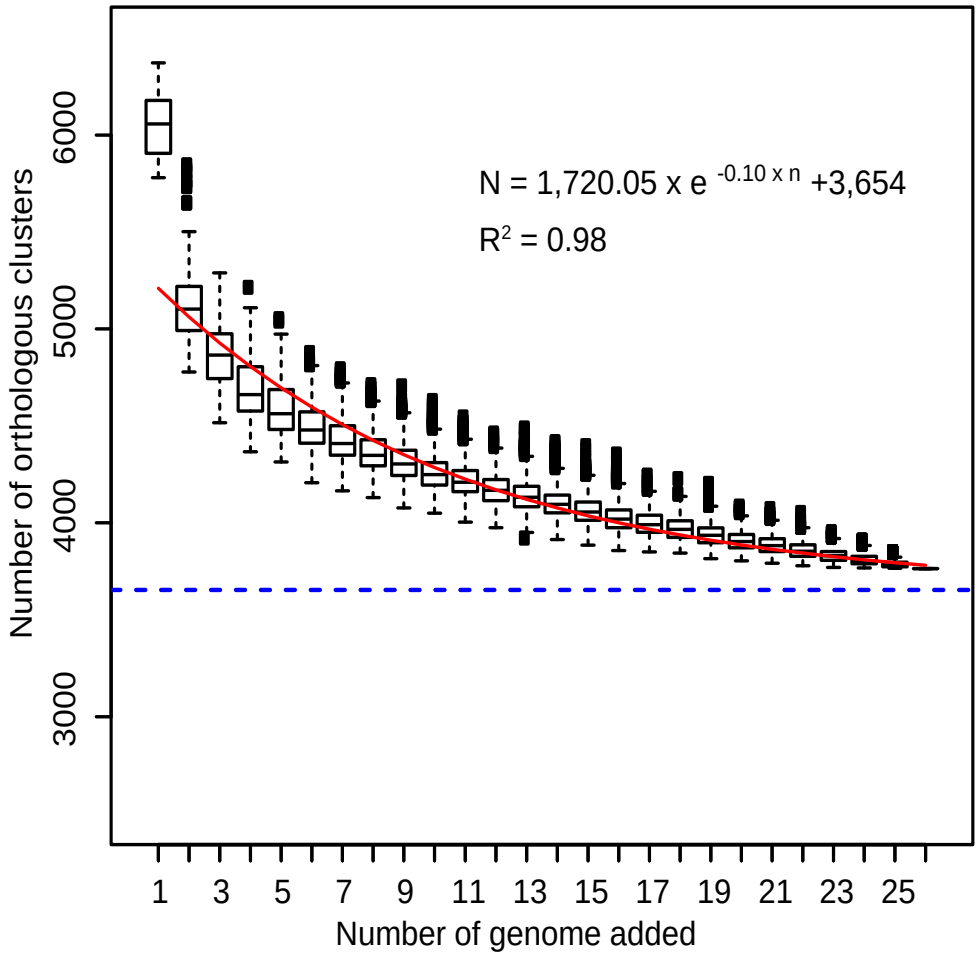


Figure 4.4: Boxplot showing the least squares fit curve of the *A. xylosoxidans* core genome. The analysis shows that the number of orthologues decreases with the number of genomes added. The red line is a least squares fit of the exponential law $N = \kappa\ e^{-n/\tau} + \Omega$, with the correlation value, $R^2 = 0.98$. This estimates that the size of the core genome of *A. xylosoxidans* is approximately half of the size of the whole genome.

4.3.5. Core genome single nucleotide polymorphisms (SNPs) and phylogeny

The relationships between strains were investigated using phylogenetic trees constructed using core genes using two different algorithms: Bayesian inference (Figure 4.5) and Maximum-likelihood (Figure 4.6). Of the 3,764 core orthologous groups, 3,737 groups were considered core genes in this study due to exactly one gene representing one genome in each orthologous group. After stripping gaps and concatenating core genes, this resulted in 2.6 Mbp.

There were similarities in the phylogenetic relationships amongst *A. xylosoxidans* in both trees. Phylogenetic analysis revealed two major clades - clade I and clade II (Figure 4.5 and 4.6) - with no specific geographical correlation. In clade I, there was a Thai-specific cluster (R2, R3, R7, R8, R9, R10 and R13), which was closely related to a Liverpool isolate, L10. There was also a close evolutionary relationship between three Liverpool isolates, which were from patients with different conditions - were L8 and L14 (non-CF) and L17 (CF) (Table 2.2). In clade II, isolates from the same patient (R1, R5 and R14) clustered together, demonstrating a close relationship with a non-CF Liverpool isolate, L1, and the reference genome, NH44784-1996. Strain NCIMB 11015 was positioned as a separate sub-clade in clade I. In terms of functional distribution of the core genome, there is no difference in the gene functions presented in these clades (Figures A2.2 and A2.3).

The Bayesian inference tree showed better statistical support than the Maximum-likelihood tree. The posterior probability value was better than the bootstrap value in the mid-rooted tree. The posterior probability values were more than 0.8 and most of them were 1.0 in the Bayesian inference tree (Figure 4.5). Most of the bootstrap supports were more than 70%, except for two values, which were 50% and 67%, in Clade I (Figure 4.6). An analysis of new genes discovered in each clade revealed that clade I seemed to be more open than clade II, as indicated by the extrapolation of new genes discovered in each clade (Appendix Figure A2.5).

In total, 105,298 positions out of 2,602,900 bp (4.05% of core genome) were identified as core genome SNPs. The proportion of polymorphic sites was, on average, 4.05% which means that 1 SNPs exists in every 25 bp. By looking at clade specifically, a total of 85,651 SNPs and 57,692 SNPs were identified within Clade I and Clade II, respectively.

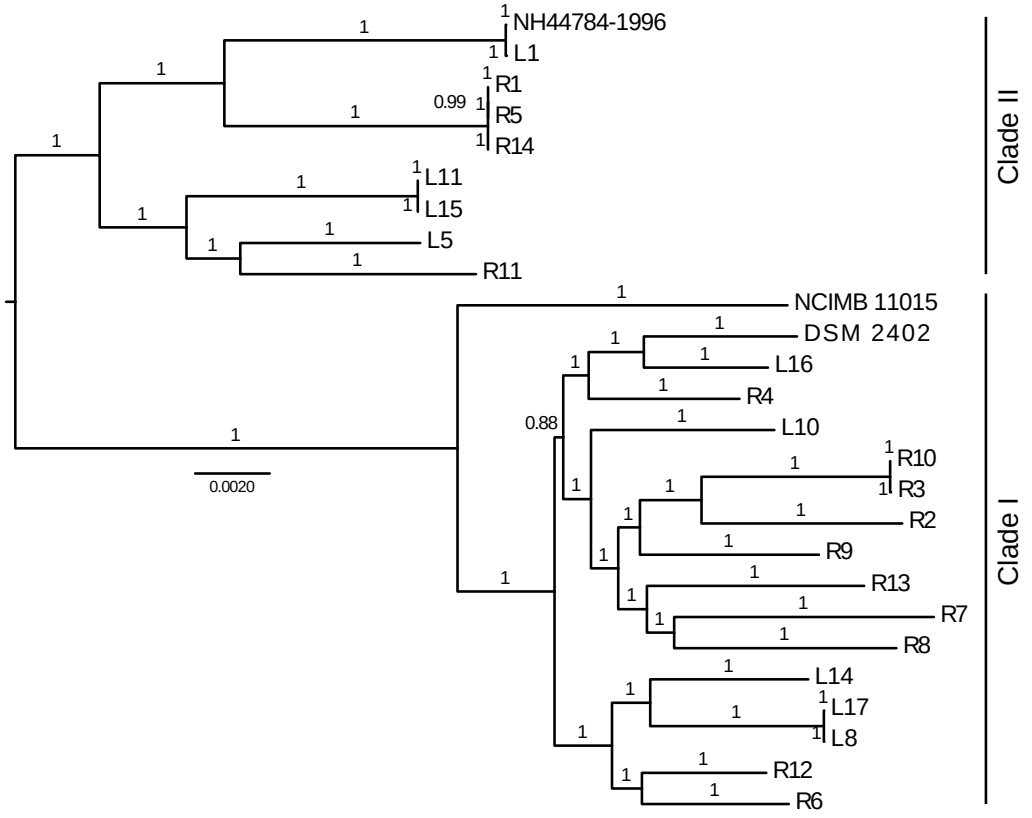


Figure 4.5: Bayesian inference mid-rooted phylogenetic tree based on 3,736 core genes from 26 genomes of *A. xylosoxidans* included in this study. Phylogenetic clades I and II are indicated. Strain nomenclature indicates the origins of clinical isolates; 'L' strains are from Liverpool and 'R' strains are from Thailand. NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; and NH44784-1996 for *A. xylosoxidans* NH44784-1996. All posterior probability values are higher than 0.8.

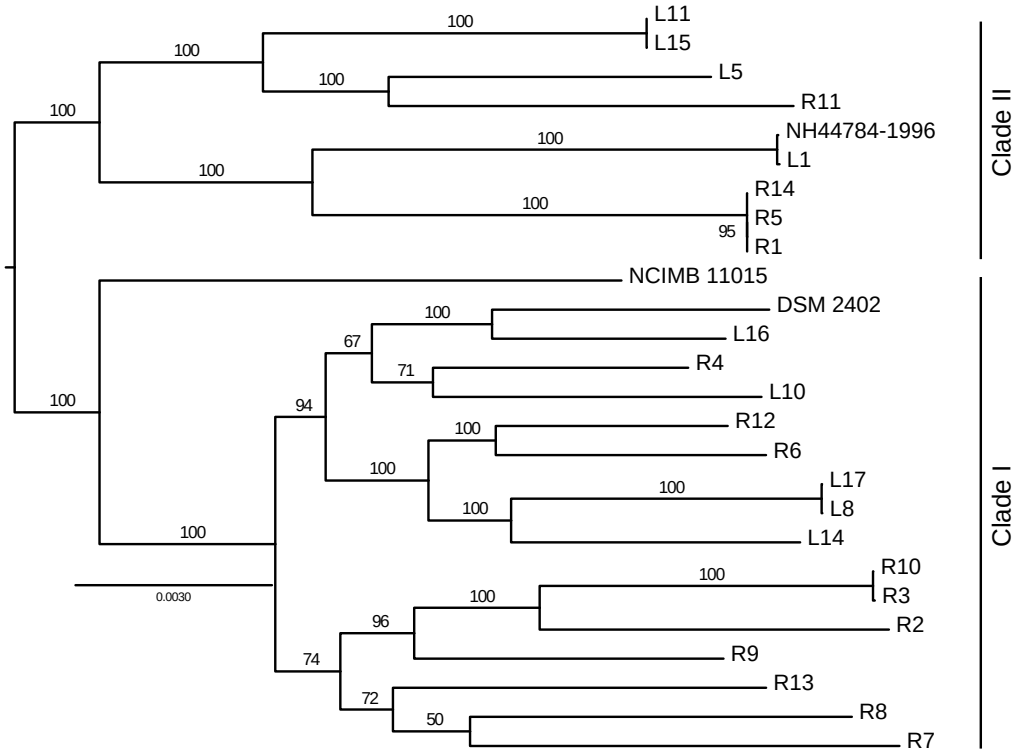


Figure 4.6: Maximum-likelihood mid-rooted phylogenetic tree based on 3,736 core genes from 26 genomes of *A. xylosoxidans* included in this study. Phylogenetic clades I and II are indicated. Strain nomenclature indicates the origins of the clinical isolates; 'L' strains are from Liverpool and 'R' strains are from Thailand. NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; and NH44784-1996 for *A. xylosoxidans* NH44784-1996. Bootstrap values more than 50% were indicated.

4.3.6. Comparative genomic analysis

For the purpose of a comparative genomic analysis, 26 genomes of *A. xylosoxidans* strains, including 25 recently sequenced genomes and the NH44784-1996 genome, were compared to one another using orthologous analysis performed by OrthoMCL software. This generated a total of 8,762 OrthoMCL groups with 3,764 groups seen in all 26 genomes. These groups were defined as 'core orthologues'. The number of genes unique to each isolates was approximately 119 genes per genome (Appendix Table A2.2). The largest of OrthoMCL's output groups was comprised of 154 genes from 26 *A. xylosoxidans* genomes, which are encoded a RND-type efflux pump inner membrane subunit. Interestingly, these efflux pumps are associated with drug resistance in *A. xylosoxidans*, which will be discussed in Chapter 5.

The determination of the functions of the core genes using Clusters of Orthologous Groups (COGs) suggested that approximately 31% of the core genes were assigned generalised and poorly-characterised functions, with most being labelled as 'hypothetical protein' (Table 4.5; Figure 4.7). Apart from genes with unknown roles, the core genes employed functions related to adaptation to the environment, including roles such as amino acid transport/metabolism (12.57%), inorganic ion transportation/metabolism (6.16%) and energy production (5.95%), and housekeeping functions such as transcription (9.65%).

For accessory orthologues (Appendix Table A2.3; Figure 4.7), 38% were assigned to functional categories. The most abundant group was associated with transcription, which made up 6.34% of the genome. The second most abundant was associated with amino acid metabolism, comprising 5.41% of the genome. Orthologues in the category 'inorganic ion transport/metabolism' represented 4.16% of the genome.

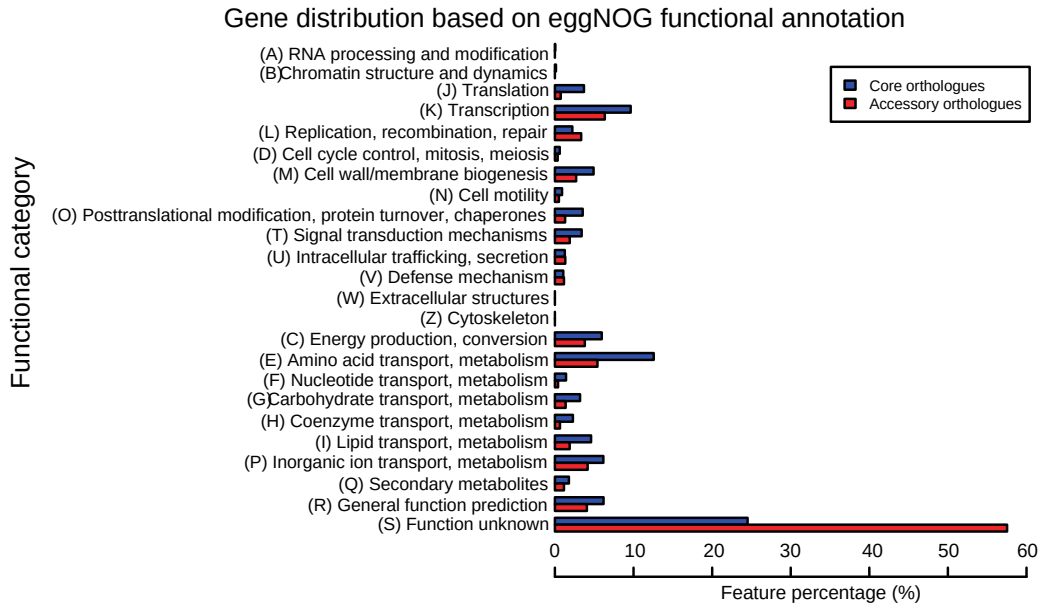


Figure 4.7: Functional annotation of core orthologous genes and accessory orthologous genes in *A. xylosoxidans.* Genes called by Prokka version 1.7.2 were entered into the eggNOG database, which annotated functions based upon Clusters of Orthologous Groups (COGs) categories.

4.3.7. Recombination

Recombination events are common in bacteria. It is one of the processes that drive genetic evolution in bacterial populations. Starting with the core genome phylogeny (Figure 4.6), a PHI test and Neighbor-Net tree implemented in SplitTree (Figure 4.8A) detected significant levels of recombination in *A. xylosoxidans*. Recombination-detecting programs, ClonalFrame, BratNextGen and ClonalFrameML, were applied to SNPs alignments extracted from the sequence alignment of the core genes. ClonalFrame did not yield any results. For BratNextGen, no clustering resulted from the analysis when default parameters were used. With the alpha-parameter set to 1, as suggested by Dr. Pekka Martinen, recombination was seen to have happened everywhere across the core genome (Figure 4.9). Recombinant fragments detected by the software were removed from the core genome alignment. The phylogenetic tree then was re-built based upon recombination-filtered alignments (Figure A2.6); however the recombination in the alignments was still detected by SplitTree and the PHI test

(Figure 4.8B). Furthermore, recombination analysis using ClonalframeML revealed a complicated 4,978 recombination events spreading through out the core genes of *A. xylosoxidans* (Appendix Figure A6.1).

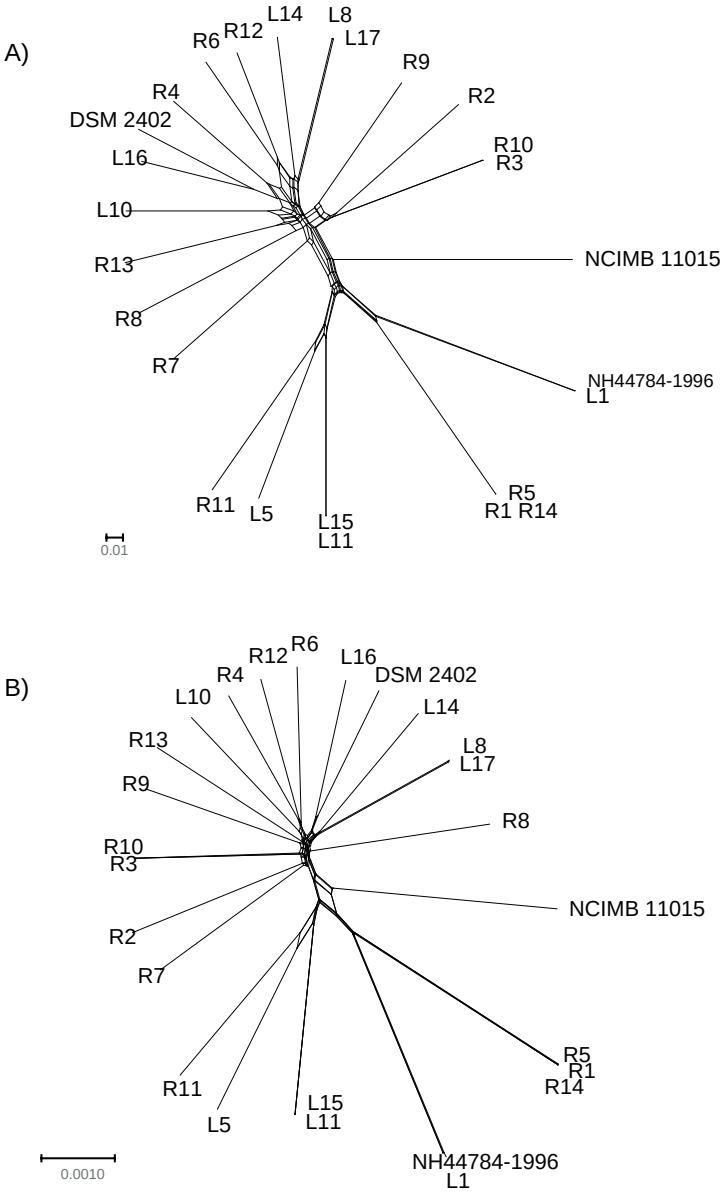


Figure 4.8: SplitTree's Neighbor-Net phylogenetic based upon SNP identified on 3,737 core genes, showing the net-like connection of core SNPs of 26 genomes of *A. xylosoxidans* A) before recombination sites were filtered out and B) after recombination sites were filtered out by BratNextGen. Strain nomenclature indicates the origins of clinical isolates: 'L' strains are clinical isolates from Liverpool; 'R' strains are clinical isolates from Thailand; NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A.*

xylosoxidans DSM 2402; and NH44784-1996 for *A. xylosoxidans* NH44784-1996.

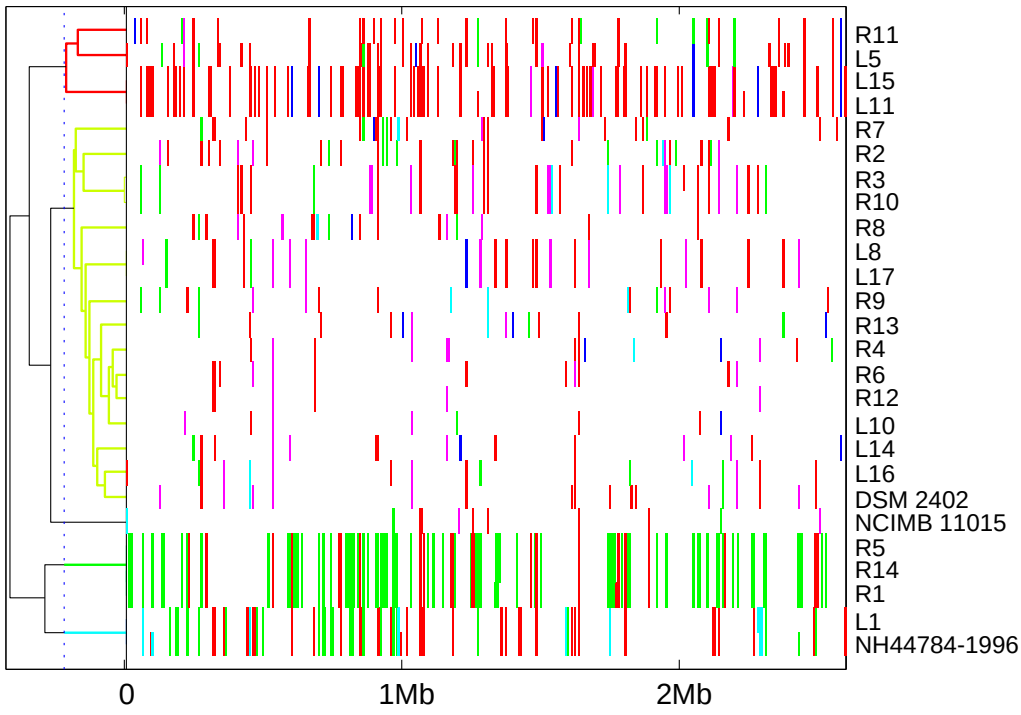


Figure 4.9: Recombination detection across the core genome of *A. xylosoxidans* performed using BratNextGen. The dendrogram on the left-hand side represents PSA (Proportion of shared ancestry), analysed by bratNextGen where alpha = 1. The colour of each block on the plot represents recombinant sequences received from strains/clusters corresponding with the colours on the dendrogram. Strain nomenclature indicates the origin of clinical isolates; 'L' strains are clinical isolates from Liverpool; 'R' strains are clinical isolates from Thailand; NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; and NH44784-1996 for *A. xylosoxidans* NH44784-1996.

To confirm that the recombination detection software works with other genomic data, BratNextGen was applied to *Bartonella hensalae* genomes. As presented in Appendix Figure A6.2, BratNextGen revealed possible recombination events in 21 *B. hensalae* genomes when an alpha-parameter of 1 was used. Of the 809 genes affected by recombination events, 51 genes were virulence-associated genes, encoding such proteins as those involved in adhesion, a hemin-binding protein and the type IV secretion system, which have been reported as major virulence genes in host-specific species/strains (Chen, 2004; Engel *et al.*, 2011).

4.3.8. Cluster analysis

To assess the phylogenomic variation and adaptation in *A. xylosoxidans*, the occurrence of accessory genes was investigated. The clustering analysis inferring the presence/absence of 4,998 orthologous groups of accessory genes clearly illustrated two major groups, namely clade A and clade B (Figure 4.10).

As shown in Figure 4.10, clade A included the majority of strains used in this study: NCIMB 11015, DSM 2402, L1, L5, L8, L10, L14, L16, L17, R1, R2, R4, R5, R6, R8, R11, R12, R13, R14 and NH44784-1996. Clade B was comprised of six strains: L11, L15, R3, R7, R9 and R10. Clade B included strains that were grouped by their country of origin - R3, R7, R9 and R10, all of Thai origins were grouped together, and L11 and L15, both of Liverpool origins, formed a single group.

Of the 4,998 gene families, 1,067 orthologues were found to be unique to clade A and 384 orthologues were unique to clade B. Using Random Forest™ to investigate lineage-differentiating genes, comparative analysis suggested that ORTHOMCL5083 and ORTHOMCL5053, which encode a putative Acetyl-CoA synthetase (ADP-forming) alpha and beta chains and major facilitator superfamily MFS 1, respectively, were statistically significant factors associated with the formation of Clade A (Figure 4.11). When the tables were analysed, these two orthologous groups were present in all members of clade B, except R13, whereas none of the clade II contained these orthologous groups.

Further investigations into clade A (Figure 4.10) revealed that the clade could be divided into three sub-clades, which were named A1, A2 and A3. Binary table inspection showed that there were no clade A3-specific orthologous groups, but there were 15 orthologous groups that existed in all strains in clade A except R1, R5 and R14. Between clade A1 and A2, Random forest™ revealed that ORTHOMCL5667, ORTHOMCL5174 and ORTHOMCL5393, each of which is annotated as 'hypothetical protein', were statistically involved in differentiation between clade A1 and clade A2 due to greatly mean decreased accuracy scores and mean decreased Gini scores (Figure 4.12). By determining an individual orthologous group, ORTHOMCL5667 was present in all of clade A2's strains and L14 (Clade A1 in Figure 4.11), whereas ORTHOMCL5174 and ORTHOMCL5393 existed in both clade A1 and A3. Therefore, only ORTHOMCL5667 was statistically reliable as a gene family that differentiated clade A2 from clade A3.

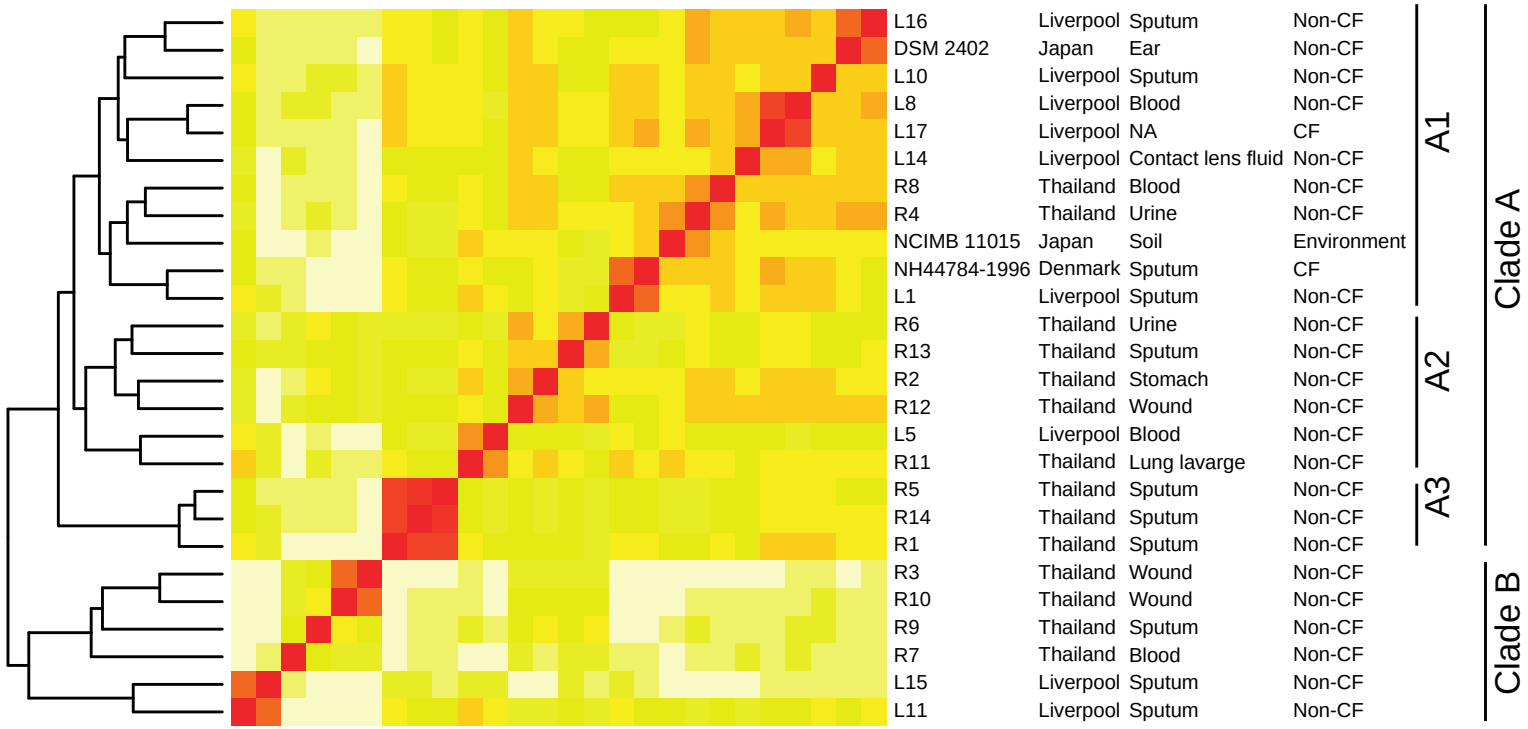


Figure 4.10: A hierarchical heatmap illustrating the clustering of *A. xylosoxidans* strains based on the occurrence of the orthologous groups of 26 *A. xylosoxidans* genomes in this study. Condition of patients (CF versus non-CF) and country of origin of isolates are included. Clade A and clade B are indicated. Strain nomenclature indicates the origin of clinical isolates: 'L' strains are clinical isolates from Liverpool; 'R' strains are clinical isolates from Thailand; NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; and NH44784-1996 for *A. xylosoxidans* NH44784-1996.

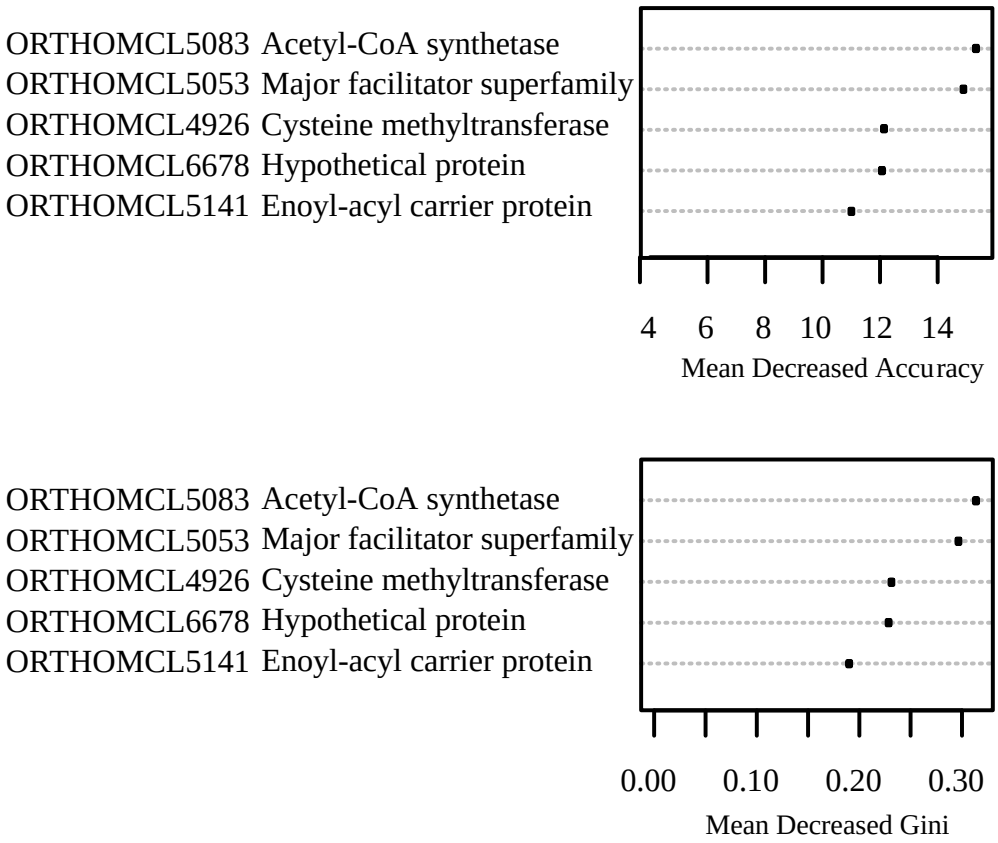


Figure 4.11: A plot of outstanding results from a Random Forest™ classification of orthologues that were associated with clade A and clade B differentiation in Figure 4.9. The analysis shows that genes in ORTHOMCL5083 and ORTHOMCL5053 were associated with the differentiation between clade A and clade B. The X-axes indicate the level of importance. The Y-axes show the list of orthologues.

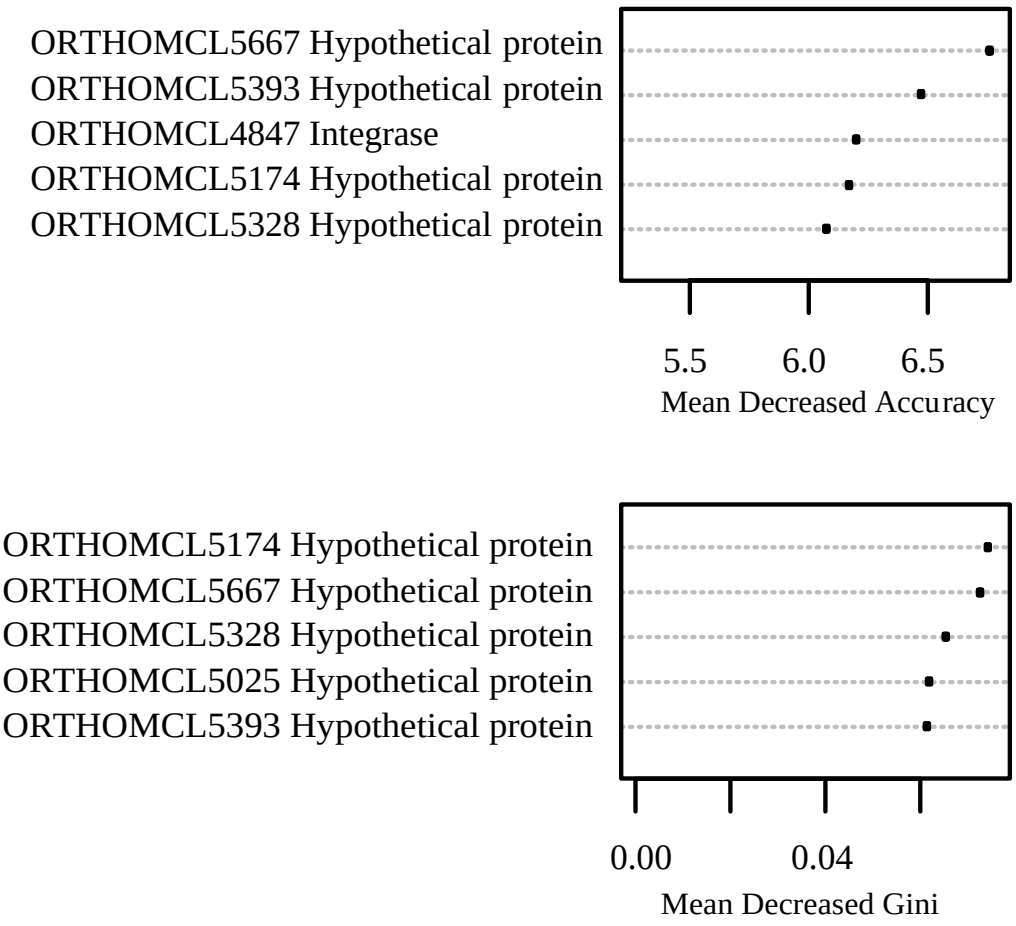


Figure 4.12: A plot of outstanding results from a Random Forest™ classification of orthologues that were associated with sub-differentiation of clade A in Figure 4.9. The analysis shows that genes in ORTHOMCL5667 were associated with the differentiation of sub-clades in clade A. The X-axes indicate the level of importance. The Y-axes show the list of orthologues.

4.3.9. Comparison of core genome phylogeny and accessory genome clustering

By comparing a core genome-derived phylogenetic tree with an accessory genes-based clustering (Figure 4.13), a tangle-gram showed the similarities in strain clusters between the core genome phylogeny and hierarchical clusters of accessory orthologues. Five pairs of genomes, which were from different sources, clustered together; they were strains NH44784-1996 and L1, L5 and R11, L8 and

L17, AX DSM 2402 and L16, and R3 and R10. Isolates from the same patient clustered together; they were R1, R5 and R14, and L11 and L15 (highlighted by a red box and red lines in Figure 4.13). The comparison of clinical isolates from the same patients, as well as the comparison of clinical isolates from a CF patient with a strain from a non-CF patient was performed. Isolates from the same patients were L11 and L15, and R1, R5 and R14. The clinical isolates from non-CF patient was L1, compared with the clinical isolates from a CF patient, strain NH44784-1999.

In the comparison of *A. xylosoxidans* genomes collected from the same patient from Liverpool, L11 and L15, 5,761 gene families were shared between these two strains and only two SNPs existed in the core genomes between them. There were 667 genes that were unique to one of the strain. For Thai strains from the same patients, R1, R5 and R14, there were 5,745 shared gene orthologues (90.5%) with three SNPs in the core genome, with 602 genes unique to one of the strains.

In total, 5,651 gene families were conserved between strains L1 and NH44784-1996, and 107 SNPs were identified in their core genomes. A total of 577 genes were unique to either of them. With the exception of genes defined as 'poorly characterised' (~63%), different genes in L1 were involved in amino acid metabolism/transportation and fatty acid metabolism/transportation, and genes that were not present in NH44784-1996 were associated with carbohydrate metabolism/transportation.

A comparison of a strain from a non-CF patient, L17, and a strain from a CF patient, L8, showed a close relationship between these strains. Conserved gene families between these strains numbered 5,734 gene families. The alignment of the core genomes revealed 40 nucleotide variations between the strains. Considering the accessory genome only, the CF strain possessed more phage-related genes and inorganic ion-related genes than L8.

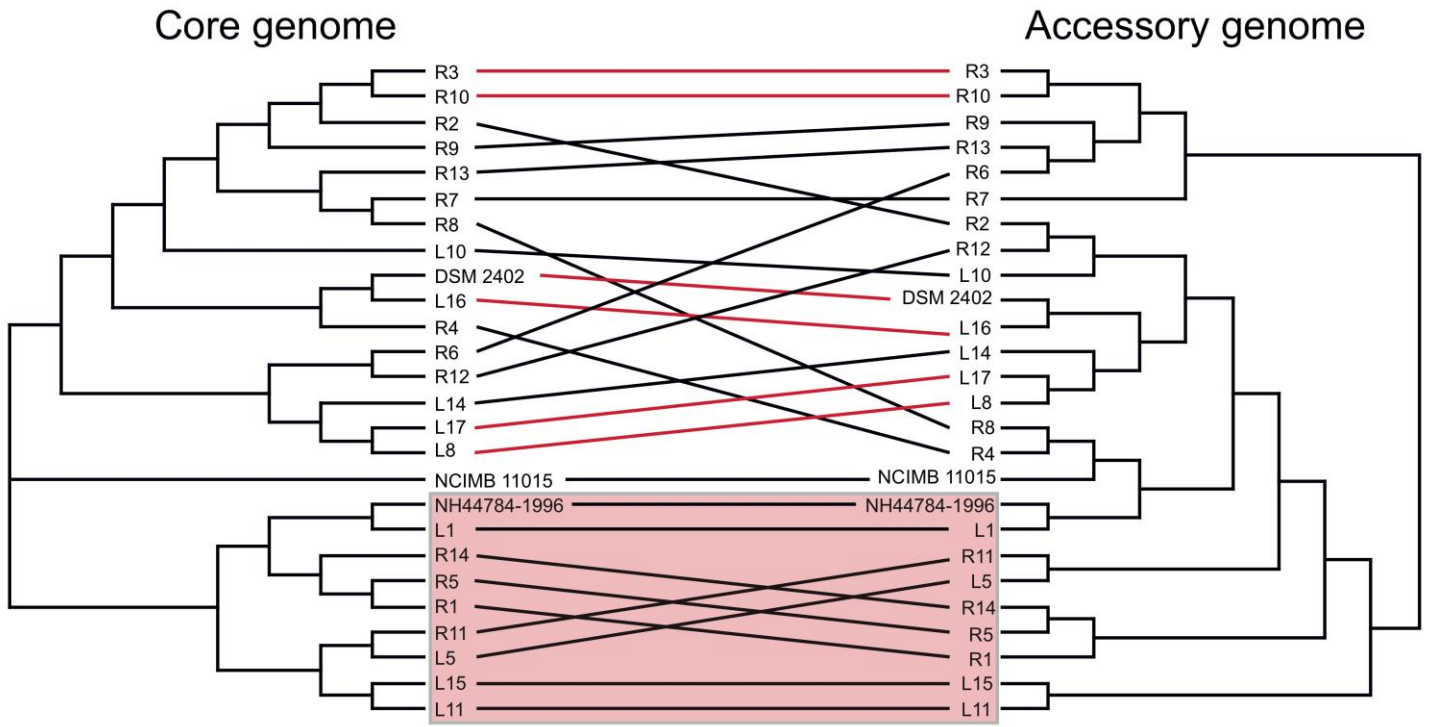


Figure 4.13: A tangle-gram tree to compare core genome phylogeny (left) and accessory genome clustering (right). The red boxes and red lines indicate genomes that are closely related to one other according to both core genome analysis and accessory genome analysis. Strain nomenclature indicates the origins of clinical isolates: 'L' strains are clinical isolates from Liverpool; 'R' strains are clinical isolates from Thailand; NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; and NH44784-1996 for *A. xylosoxidans* NH44784-1996.

4.4. Discussion

A. xylosoxidans is an emerging pathogen causing infections in Cystic Fibrosis patients (Dunne & Maisch, 1995; Krzewinski *et al.*, 2001; Liu *et al.*, 2002; Ferroni *et al.*, 2002; Gibson, Burns & Ramsey, 2003; Van Daele *et al.*, 2005; Rønne Hansen *et al.*, 2006; De Baets *et al.*, 2007; Spicuzza *et al.*, 2009; Pereira *et al.*, 2011; Ciofu, Hansen & Høiby, 2013; Mahenthiralingam, 2014). *A. xylosoxidans* can also cause problems in non-CF infections (Yabuuchi & Oyama, 1971; Holmes, Snell & Lapage, 1977), including, opportunistic infection (Igra-Siegman, Chmel & Cobbs, 1980; Giacoia, 1990; Reverdy *et al.*, 1984; Arroyo *et al.*, 1987; Gómez-Cerezo *et al.*, 2003; Tena *et al.*, 2005). So far, there have been a number of studies focusing on different kinds of infection caused by *A. xylosoxidans* (Adam *et al.*, 2013; Tatro *et al.*, 2013; Lee *et al.*, 2014; Tena *et al.*, 2014; Otta *et al.*, 2014). Despite the fact that previous works demonstrated comparative analysis of the members of genus *Achromobacter* (Li *et al.*, 2013) and antibiotic resistance genes in *A. xylosoxidans* (Hu *et al.*, 2015), a comprehensively comparative analysis of *A. xylosoxidans* genomes has not yet been performed. *A. xylosoxidans* from a teaching hospital in Thailand and clinical sites in Liverpool have been used here in a comparative genomic analysis of the species to obtain comprehensive information about the genetic diversity of *A. xylosoxidans* from different areas, and obtain a better understanding of the *A. xylosoxidans* pan-genome.

4.4.1. Phylogenetic relationships reveal global characteristics of *A. xylosoxidans*

This is the first study to conduct comparative genomic analysis of *A. xylosoxidans* strains from different locations. A phylogenetic analysis of *A. xylosoxidans* (Figure 4.6 and Figure 4.7) suggests a global and diverse population structure of *A. xylosoxidans*, as implied by the absence of geographical location-specific clades. Heatmap clustering based on accessory genomes also presents non-specific clades amongst these genomes. Interestingly, some genomes from different locations are paired together, including those of

strains NH44784-1996 and L1, L5 and R11, L8 and L17, and DSM 2402 and L16. This grouping of genomes by core genome phylogeny and disposable genomes is also in agreement with strain typing performed using RAPD (Figure 3.6) and MLST (Figure 3.7). These finding suggests that geographical feature of does not reflect genomic structure of *A. xylosoxidans*. This is not in agreement with other species. Genomic analyses in human pathigens, such as *V. cholerae* and *C. fetus*, illustrate that geographical identity and niche specificity of the pathogens can be influenced by genomic structure, mobile genetic elements in particular (Dutilh *et al.*, 2014; Kienesberger *et al.*, 2014).

Also, this can suggest the global distribution of *A. xylosoxidans* in the environments (Bador *et al.*, 2013). It is noteworthy that compromised patients will probably be infected by bacteria from the environment. Moreover, the complication regarding the epidemiology of the species would be seen during outbreak, as there is no geographical specificity. The global diversity characteristic is also present in other pathogens. For example, similar frequencies of clonal complex of *S. aureus* were seen in French and Algerian isolates. This similarity was probably driven by the migration of population between these two countries (Ruimy *et al.*, 2009). Considering *A. xylosoxidans*, the global diversity of the isolates can be occurred by the migration of population between Thailand and the U.K. However, the greater number and the more diversity of sites of *A. xylosoxidans* isolates are required for further investigation.

4.4.2. The possibility of genetic recombination in *A. xylosoxidans*

Recombination is a key driver that causes a huge leap of evolution in bacteria because bacteria can exchange genetic fragments with one other. This potentially leads to a significant change in the bacteria, enabling more rapid evolution than can be achieved by the small changes that arise due to single mutations during replication (Boinett & Cain, 2014). Also, recombination can be associated with important phenotypes such as antibiotic resistance if the recombination involves transfer of drug resistance genes (Hanage *et al.*, 2009; Croucher *et al.*, 2014).

In a previous study of six *Achromobacter* genomes (Li *et al.*, 2013), a low rate of recombination between different species of genus *Achromobacter* was detected; however, this could have been underestimated due to the diversity of genomes used in the study. Recombination analysis is a problem in this study. Analysis using a Neighbor-Net-derived phylogenetic tree and a PHI test in SplitTree revealed significant levels of recombination signals within the core genome alignment. However, ClonalFrame could not reveal recombination segments within the core genome. It is important to note that ClonalFrame is suitable for alignments of small regions/genes such as MLST analysis (Didelot & Falush, 2007), rather than large data sets, such as the core genome's alignment. BratNextGen is a program that is able to detect and identify recombination in various bacteria including *Streptococcus pneumoniae* (Marttinen *et al.*, 2012; Croucher *et al.*, 2013; Chewapreecha *et al.*, 2014), *Staphylococcus aureus* (Castillo-Ramírez *et al.*, 2012), *Escherichia coli* (McNally *et al.*, 2013) and *Enterococcus faecium* (de Been *et al.*, 2013). In this study, the software did detect significant recombined fragments but it could not resolve recombination hotspots across 26 *A. xylosoxidans* genomes. Howell *et al.* (2014) experienced similar limitations when performing a recombination analysis using BratNextGen with *Haemophilus parasuis* genomes. Taking all of the above into consideration, it seems that there was evidence of recombination in the *A. xylosoxidans* genomes; however, genetic regions affected by recombination were not clearly identified by softwares used in this study (BratNextGen and ClonalFrame ML). Another recombination software (Gubbins) will be considered. Taking all above into account, the selection of the diversity of the samples (genomes) and using multiple softwares are important for recombination analysis in bacteria.

Moreover, it has an influence on virulence and drug resistance. Together with this evidence and these implications, it appears that recombination guides *A. xylosoxidans* to acquired and antimicrobial drug-resistant phenotype, which will be discussed in Chapter 5.

4.4.3. Pan-genome characteristic correlates with receivable characteristic of *A. xylosoxidans*

Li *et al.* (2013) revealed that pan-genome of six genomes of members of genus *Achromobacter* is opened; however, they did not focus on *A. xylosoxidans*. The prediction of the pan-genome was in agreement with the open genome characteristic. In addition, with Heaps' law, the model of new genes acquired supported the open-genome, which implies continued exchange of genes between the gene pool of the species and environment (Tettelin *et al.*, 2008). Interestingly, pairwise comparison of clinical isolates obtained from the same patients revealed the difference between the isolates. Considering core-genome phylogeny and the heatmap of the existence of accessory genome, L11 was closely related to L15, and R1, R5 and R14 were closely related to one another. However, there was approximately 10% difference between them. This can be also explained by the open pan-genome characteristic of the species (Figure 4.2 and Figure 4.3). This is in agreement with the evidence of genetic recombination. Examples of factors that drives genetic exchange in bacteria are mobile gene element transfer (Hu *et al.*, 2015; Chen *et al.*, 2014) and bacteriophage-related gene transfer (Wittmann *et al.*, 2014). Previous studies report that genetic exchange occurring in *A. xylosoxidans* contributes to antibiotic resistance, such as genes encoding carbapenemase (Yamamoto *et al.*, 2012; Hu *et al.*, 2015). This is a problem that needs to be concerned in *A. xylosoxidans* infection.

When considereing the genomic characteristics and pathogenicity of a species, pathogens with a closed-genome are assumed to be real pathogens because they do not need any virulence genes from other species (Rouli *et al.*, 2014). Pathogens with open-genome, on the other hand, are likely to adapt from commensal strains or non-pathogenic bacteria to survive in the human body (Table 4.5). *Bacillus anthracis* is an example of a pathogen with a closed-genome (Rouli *et al.*, 2014). Its core genome covers approximately 99% of the whole genome's size. This study's pan-genome analysis considers *B. anthracis* an absolute pathogen due to the stability of the genome. Here, with its open-genome, *A. xylosoxidans* would have the ability to exchange its genetic materials with surrounding pathogens. In other words, *A. xylosoxidans* can receive

virulence genes or antibiotic resistance genes, and it can donate those genes to surrounding organisms. This suggests that this organism could have a major impact upon public health in the future by acting as a genetic reservoir.

Table 4.5: Examples of pan-genome analyses on different organisms

Pan-genome	Organism
Opened	*Streptococcus agalactiae* (Tettelin *et al.*, 2005) *Haemophilus influenzae* (Hogg *et al.*, 2007) *Escherichia coli* (Rasko *et al.*, 2008) *Streptococcus pneumoniae* (Donati *et al.*, 2010) *Enterococcus faecium* (Qin *et al.*, 2012) *Streptococcus mutans* (Song *et al.*, 2013)
Closed	*Salmonella* (Jacobsen *et al.*, 2011) *Bacillus antracis* (Rouli *et al.*, 2014)

4.4.4. Antibiotic resistance associated efflux pumps are abundant as the largest OrthoMCL's output in *A. xylosoxidans*

One of the most concerning problems caused by *A. xylosoxidans* infections is the species' multidrug resistance. Whilst antimicrobial resistance is an important factor for clinicians when treatment is needed, an in-depth study of association between antimicrobial resistance phenotypes and corresponding genes is, perhaps, useful in studying resistance mechanisms to particular drugs. Several studies reveal that multidrug resistance in bacteria can possibly be regulated by horizontal gene transfer. However, *A. xylosoxidans* strains exhibit natural drug resistance due to their intrinsic resistance genes, for examples, OXA-114 beta-lactamase (Doi *et al.*, 2008) and RND-type efflux pumps (Bador *et al.*, 2011, 2013). In a similar way to previous studies where members of the genus *Achromobacter*, especially *A. xylosoxidans*, were found to possess multiple numbers of efflux pumps, the analysis of orthologous genes here showed that *A. xylosoxidans* strains possess high numbers of genes encoding efflux pumps within their genomes. The RND-type efflux pump ranks as the largest gene group amongst efflux pumps and amongst all genes in *A. xylosoxidans* (Hu *et al.*,

2015). This is in accordance with this study, which showed that an orthologous groups consisting of the inner membrane subunit of the RND-type efflux pump – CmeB - is the largest protein group across 26 genomes of *A. xylosoxidans*. The RND-type efflux system is comprised of three parts: the inner membrane protein, the outer membrane protein, and an accessory protein or membrane fusion protein (Nikaido, 2011). In previous studies, Bador *et al.* (2011, 2013) indicated two efflux systems, AxyABM and AxyXY-OprZ, which are associated with antibiotic resistance in *Achromobacter* species. In the same manner, this analysis addressed two RND-type transport systems, which are homologous to AxyABM and AxyXY-OprZ. Further analysis and detail of antibiotic resistance in *A. xylosoxidans* will be further discussed in Chapter 5.

4.5. Conclusions and future work

This chapter has demonstrated a comparative analysis of *A. xylosoxidans* genomes from different sources. In this study, next-generation sequencing technologies, combined with bioinformatics tools, were used to investigate the population structure and genomic characteristics of *A. xylosoxidans*.

Analysis of 25 newly sequenced genomes of *A. xylosoxidans* together with a publicly available reference genome, *A. xylosoxidans* NH44784-1996, showed a global relationship between strains. Isolates from different geographical sources (Thailand and Liverpool) did not cluster by either the core genome or the accessory genome, agreeing with analysis based on MLST typing (Figure 3.7) and suggesting that there are no distinct geographical associations with particular genotypes. Nonetheless, the limitations of this study are the number of samples collected and the range of sampling locations. Further study will benefit from expanding the range of isolates, in terms of number and location, in order to achieve a comparative genomic analysis of *A. xylosoxidans*.

The study of the core-genome suggested that 3,763 gene families (approximately half of the genome's size) are conserved within the species. The analysis of new genes obtained, together with the pan-genome, allows for this *Achromobacter* species to be described as having an open-genome, which indicates the possibility of genetic exchange, such as recombination and mobile genetic elements, within the species or with different species. Furthermore, *in silico* analysis of recombination across the genomes of 26 strains of *A. xylosoxidans* using SNPs from the core gene alignment indicated that a recombination event occurred in these *A. xylosoxidans* genomes. Nevertheless, affected recombination regions were not successfully detected due to the large diversity of the genomes. This, therefore, guides us to perform additional work on recombination analysis within the core genome's clades or within regions from whence isolates were obtained. Moreover, comparative genomic analysis focused on antibiotic resistance and pathogenicity of *A. xylosoxidans* is required.

Chapter 5

The genetics and phenotype of antibiotic resistance in *Achromobacter xylosoxidans*

5.1. Introduction

5.1.1. General overview of antibiotics

Antibiotics improve life expectancy and increase survival rates of patients, notably reducing the number of patients who die as a result of an infection following an operation. Dating back to 1941, when penicillin was first administered, antibiotics have been used to treat patients infected with bacteria, whilst research into antibiotic discovery and development has continued. Antibacterial agents work via four main mechanisms (Katzung, Masters & Trevor, 2009):

1. Interrupting and inhibiting cell wall production
2. Altering cell membrane permeability
3. Interfering with protein synthesis
4. Interfering with nucleic acid synthesis

The first mode of action is that the agent interrupts bacterial cell wall production by binding to the protein that elongates the cell wall structure or by directly binding to cell wall components. Without a cell wall, bacteria will become shrivelled in hypertonic solutions or be lysed in hypotonic solutions. Agents acting in this manner include the beta-lactams and glycopeptides. Beta-lactam is the most famous group of antibiotics, widely used across the world due to the spectrum of drugs available. The chemical structure of the beta-lactams mimics the structure of D-alanyl-D-alanine (Figure 5.1). Beta-lactam agents inhibit cell

wall synthesis by competitively binding to the D-alanyl-D-alanine transpeptidase enzyme or penicillin-binding proteins (Kong, Schneper & Mathee, 2010). This results in no cross-linking between murein polymers, thus weakening the cell wall. Examples of groups of drugs in this category are penicillin, cephalosporin and carbapenem.

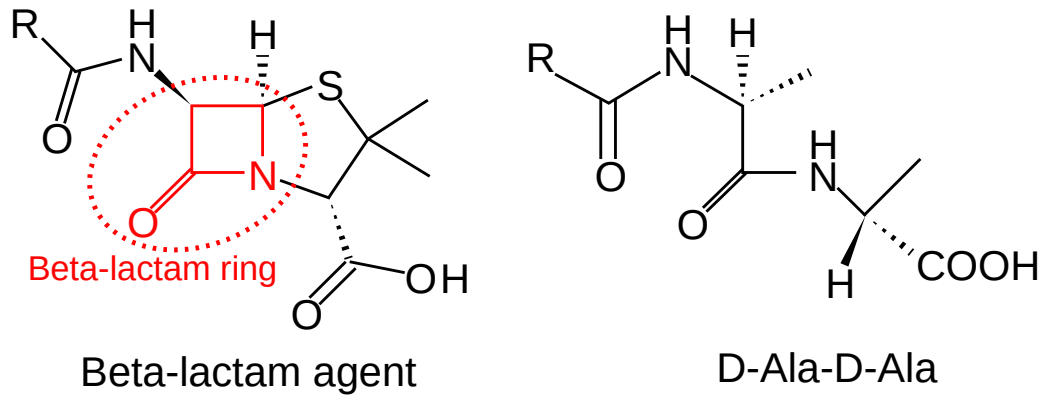


Figure 5.1: Chemical structures of penicillin (a beta-lactam antibiotic) and D-alanyl-D-alanine (cell wall component). This demonstrates the similarity of –CO-N- group in both chemicals. The beta-lactam ring is highlighted in red.

The synthesis of bacterial cell walls is not interrupted only by the inhibition of the enzyme that processes the biosynthesis but also by interference with biosynthesis molecules. Glycopeptide antibiotics act to inhibit the cross-links between murein molecules by attaching directly to the D-alanyl-D-alanine moiety of murein molecules. This attachment results in the cell's inability to tie the cell wall. Vancomycin is an example of a glycopeptide antibiotic.

The second mechanism of action is to interrupt and change membrane permeability so as to allow water or other antibiotics to get into, and kill bacteria. Drugs in this group should contain a hydrophobic group as well as a hydrophilic group that makes the chemicals insert themselves within the cell membrane and disrupt the membrane's structure and permeability. In other words, the agents behave as a detergent. Polymyxin and colistin exemplify drugs employing this mode of action.

The third mode of action seen in drugs is to interrupt or to inhibit cellular protein biosynthesis. Proteins are the main components that drive cellular functions and keep cells alive. Therefore, inhibiting protein biosynthesis can stop bacteria from growing and/or can kill them. Chemicals in this category include aminoglycoside, chloramphenicol, tetracycline and macrolide. These agents take part in protein synthesis inhibition by (1) preventing the formation of an initiation complex and inducing incorrect translation (aminoglycoside), (2) preventing the sitting of amino-acyl tRNA (tetracycline), (3) preventing the formation of a peptide bond (chloramphenicol) and (4) inhibiting translocation of the ribosome (macrolide) (Figure 5.2).

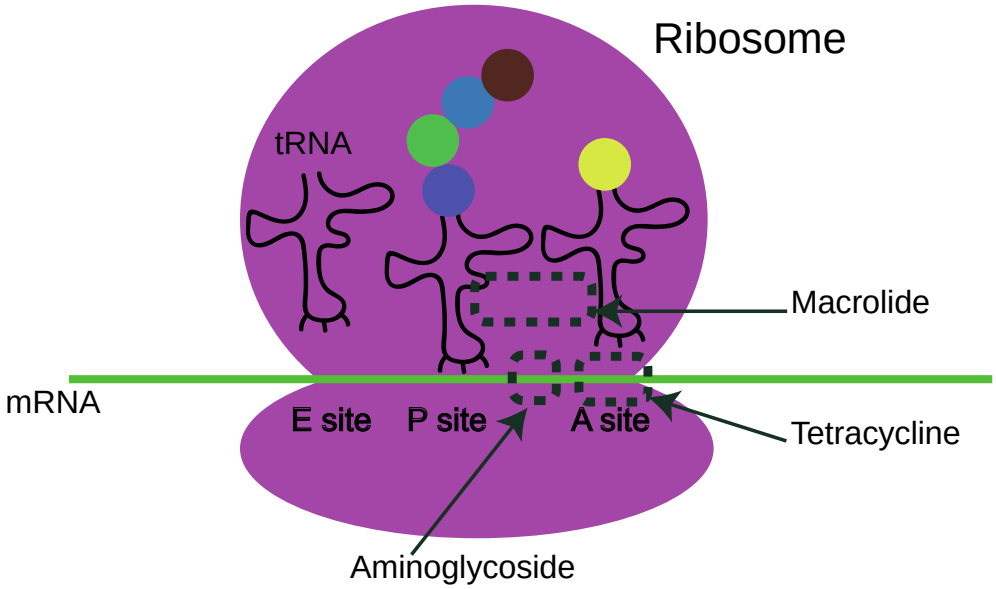


Figure 5.2: A diagram to show an mRNA-ribosome complex and target sites for multiple antibiotics. Dashed boxes illustrate the binding sites at which the indicated antibiotics act.

The fourth way that antibiotics work is to interfere with nucleic acid synthesis, which ultimately halts bacterial division and reproduction. There are two major drug targets related to nucleic acid synthesis: the folate biosynthesis pathway and the DNA replication process.

Tetrahydrofolate is essential to nucleotide biosynthesis and methionine biosynthesis in bacteria. Briefly, tetrahydrofolate acts as a methyl group carrier and provides methyl groups for the biochemical synthesis of many molecules

including purines, pyrimidines and methionine (Figure 5.3). Dihydropteroate synthase and dihydrofolate reductase are key enzymes in folate metabolism, which does not take place in humans, and they are inhibited by sulfamethoxazole and trimethoprim, respectively (Figure 5.3).

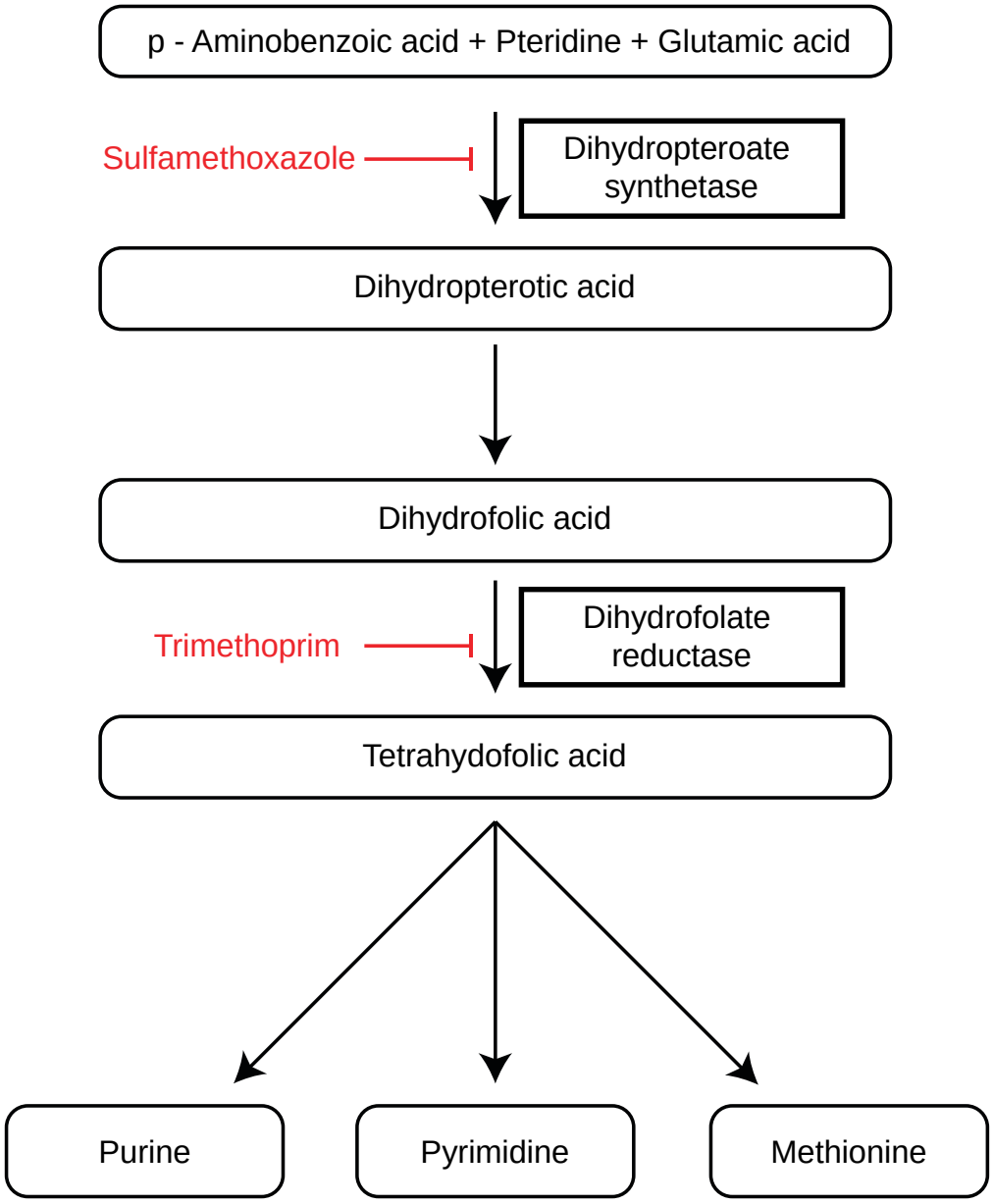


Figure 5.3: Tetrahydrofolate biosynthesis pathway in prokaryotic organisms. The diagram shows two enzymes that are inhibited by the two antibiotics: sulfamethoxazole and trimethoprim. The inhibition of these enzymes leads to a shortage of substrates for nucleotide and methionine biosynthesis in prokaryotic organisms.

DNA replication is another target for antimicrobial agents. During DNA replication, the movement of the replication induces the formation of positive-supercoils downstream to the replication site. However, negative-supercoils are preferred during the DNA replication. The topology of DNA is changed from positive-supercoils into negative-supercoils by DNA topoisomerase (Witz & Stasiak, 2010). The reaction happens at the binding pocket of topoisomerase and involves cutting, unwinding and rejoining of the DNA strands, allowing DNA replication to continue. The function of DNA topoisomerase can be inhibited by agents in category fluoroquinolones (Drlica, 1999). Fluoroquinolone antibiotics, including ciprofloxacin and levofloxacin, target the enzyme by binding to the pocket. As a result, DNA replication is interrupted by improper topoisomer of the DNA supercoils, leading to the inhibition of bacterial growth.

5.1.2. Antibiotic resistance in Gram-negative pathogenic bacteria

Clinical use of antibiotics dates back to the time when penicillin was discovered by Sir Alexander Fleming and was first synthesised by Sir Howard Florey. Since then, natural products, semi-synthetic substances and synthetic substances have been developed and trialled. Antibiotics have become a major weapon in dealing with bacterial infections. However, pathogens have become less susceptible and/or resistant to antibiotic agents. An increase in an antibiotic's use is a driving force in causing the emergence of drug-resistant bacteria. The Infectious Diseases Society of America has monitored six resistant bacteria referred to as 'ESKAPE': *E. faecium, S. aureus, K. pneumoniae, A. baumanii, P. aeruginosa* and *Enterobacter* species. Although *S. aureus* has attracted attention from researchers and clinicians around the world with the eagerness to invest money in research into methicillin-resistant *S.aureus* (Boucher *et al.*, 2009), other pathogens, in particular Gram-negative bacteria, are more problematic, especially in hospitals. WHO Antimicrobial resistance global surveillance 2014 reports that antibiotic resistance was noticed in several common pathogens including *E. coli, S. pneumoniae* and *Salmonella* species (http://www.who.int/drugresistance/documents/surveillancereport/en/).

In Gram-negative bacteria, there are four main mechanisms that drive drug resistance (Torres, Villegas & Quinn, 2007):

1. Enzyme modification that involves a chemical reaction that changes the molecular structure of the chemical, resulting in loss of drug function
2. Efflux transport system that pumps substances, such as antibiotics, out of organisms
3. Membrane permeability change due to the mutation of porin protein
4. Mutated drug target, which results in inability to bind to the target

Enzymatic modification is the process by which an enzyme modifies the structure of an antibiotic through various chemical reactions, for example, hydrolysis and acetylation. A good example of a resistance-associated hydrolytic enzyme is beta-lactamase. It has become well known since it was first described in environmental isolates (Abraham & Chain). Interestingly, in Gram-negative bacteria, it is a major factor that contributes towards beta-lactam resistance because it resides in the periplasmic space, which the agent reaches before moving into the cytoplasm. Currently, beta-lactamases are grouped into 4 categories – A, B, C and D - based on similarities in their amino acid sequences (Bush, 2010). Acetylation is another mechanism that can alter and inactivate antibiotics, involving the addition of an acetyl group to an antibiotic that results in the inactivation of the agent. A noticeable example of an acetylating enzyme is aminoglycoside acetyltransferase, which adds an acetyl group to aminoglycoside resulting in an inactivated agent (Dowding, 1979).

Pumping out is another mechanism that decreases the accumulation of foreign substances by forcing them to leave the cell. Efflux pumps contribute towards this role. In pathogenic bacteria, five families of antibiotic resistance-associated efflux pumps are defined: major facilitator (MF), multidrug and toxic efflux (MATE), resistance nodulation division (RND), small multidrug resistance (SMR) and ATP-binding cassette (ABC) (Webber, 2002; Sun, Deng & Yan, 2014). Normally, pumping mechanisms alone rarely have a significant effect

upon MIC; however, an overexpression of, particularly, intrinsic efflux pumps may bestow upon a multidrug resistance mechanism (Webber, 2002).

Cellular permeability is regulated by the presence of porins, membrane proteins that act as filters in cells. Many antibiotics penetrate cells via these channels. Some antibiotics have their own specific channels; for example, carbapenem enters cells via the OprD porin only. The mutation of these membrane proteins causes decreased influx of antibiotics into bacterial cells. Moreover, reduced numbers of porins and reduced permeability of porins indicate resistance in strains as seen, in HMP-AB in *A. baumannii* (Gribun *et al.*, 2003).

The last possible antibiotic resistance mechanism is the modification of drug targets. This modification is due to individual mutations or mutated genes acquired from other bacteria. Resistance-causing mutations can arise in any drug target inside a bacterial cell including, penicillin-binding protein, dihydropteroate synthase and topoisomerase. The mutation of a penicillin-binding protein is the most well-known example of target modification due to the evolution of methicillin-resistant *S. aureus* (Hartman & Tomasz, 1984). The mutation of drug target can be transferred from animal pathogens to human pathogens. For example, fluoroquinolone resistant *Campylobacter* induced by extensive use of fluoroquinolone in poultry was transferred from chicken to human, leading to fluoroquinolone resistant *Camplylobacter* in human (Endtz *et al.*, 1991).

5.1.3. The transfer of antibiotic resistance via integron

The exchange of genes allows bacteria to evolve and to adapt to certain environments. Even within the same species, bacteria can behave differently on account of the expression of obtained genetic elements (Ochman, Lawrence & Groisman, 2000). Horrizontal Gene Transfer (HGT), also known as Lateral Gene Transfer, is a form of genetic transfer that does not happen via reproduction, but occurs as genetic elements move from one bacterium to another bacterium (Stokes & Gillings, 2011). HGT has a remarkable impact on the evolution and adaptation of bacteria because it increases the chance of a phenotype change

more than the accumulation of single mutations does. Antibiotic resistance is a phenotype that is influenced by genetic exchange. As a result of exchange, different antibiotic susceptible profile can be present in the same species (Maiden, 1998).

A specific type of genetic element transfer involves integrons. The integron is a genetic element that enhances gene transfer using its internal cassette. In general, the structure of an integron contains three parts: the integrase gene, a recombination site, and a gene cassette (Figure 5.4).

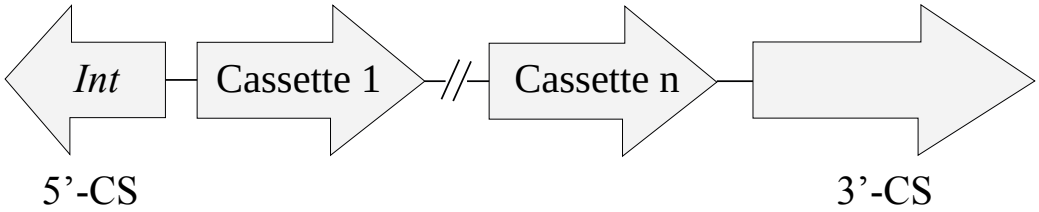


Figure 5.4: The general structure of integron. A diagram presents the structure of integron. The first part is 5' conserved sequence coding integrase. This is the class determinant of the integron, depending on the class of integrase (class 1 – class 5). The second part consists of gene cassettes obtained from the environment, catalysed by integrase. The last part is 3' conserved sequence, which is varied in different integrons.

A single integron can have more than one gene cassette (Stalder *et al.*, 2012). There are five classes of integron defined, based on the homology of the amino acid sequences of integrase genes. Class 1 integrons are the best-known type due to their being extensive studied. This integron class is evolutionarily related to Tn402. Interestingly, the integron is widely associated with antibiotic resistance transfer in bacteria and its sequence is highly conserved (Stalder *et al.*, 2012). The conserved sequence at the 5' end (5'-CS) of the element is used to determine the presence of the integron in bacteria (Gillings *et al.*, 2009). The structure of 3'-CS is comprised of a quaternary ammonium compound transporter gene (*qacEΔ*), and sulfonamide-resistance dihydropteroate synthase (*sul1*). The former is associated with quaternary ammonium compound resistance and the latter is associated with sulfonamide resistance (Paulsen *et al.*, 1993). As this type of integron, containing *qacEΔ* and *sul1*, is commonly found in clinical samples,

some authors call it ‘clinical class 1 integron’ (Gillings *et al.*, 2009). The second class of integron, class 2 integrons, is believed to be in association with Tn7. The conserved sequence of class 2 integrons is comprised of antibiotic resistance genes such as trimethoprim-resistant dihydrofolate reductase (*dfrA1*) and streptomycin-resistant aminoglycoside adenyltransferase (*aadA1*) (Hansson *et al.*, 2002). The third class of integron has been recognised not to have an impact on clinical microbiology (Rowe-Magnus & Mazel, 2002; Stalder *et al.*, 2012). However, recent studies reveal that class 3 integrons have a role in the dispersion of antibiotic resistance (Barraud *et al.*, 2013; Kargar *et al.*, 2014). Another two classes of integron have not been frequently identified, however the literatures suggest that they are associated with the development of antibiotic resistance in genus *Vibrio* (Sørum, Roberts & Crosa, 1992; Hochhut *et al.*, 2001).

There are a number of studies reporting that integrons, especially class 1 integrons, carry antibiotic resistance genes. Class 1 integrons are disseminating in the environment. Furthermore, the class 1 integrons that do not carry antibiotic resistance gene cassettes are often found in environmental bacteria, Betaproteobacteria in particular (Gillings *et al.*, 2008). In addition to the aforementioned 3’-CS that contains antibiotic resistance genes, the integron can carry resistance genes on its gene cassette. The gene cassette is able to carry various kinds of antibiotic resistance genes and most of them are aminoglycoside resistance-related genes and beta-lactamase genes (Rowe-Magnus & Mazel, 2002; Partridge *et al.*, 2009; Zhao & Hu, 2011).

5.1.4. Antibiotic resistance in *Achromobacter xylosoxidans*

The colonisation of *A. xylosoxidans* in immune-compromised patients is notable as its antibiotic resistance causes difficulties in using antibiotics to treat those patients. Since *A. xylosoxidans* was identified by Yabuuchi and Ohyama (Yabuuchi & Oyama, 1971), it has been recognised as a penicillin and aminoglycoside resistant pathogen. Reduced susceptibility to cephalosporin in *A. xylosoxidans* has also been reported in clinical isolates (Pien & Higa, 1978; Shigeta *et al.*, 1978). The molecular mechanism of cephalosporin and

aminoglycoside resistance is regulated by resistance-nodulation-cell division (RND)-type efflux pump systems. The AxyABM pump is associated with cephalosporin and fluoroquinolone resistance (Bador *et al.*, 2011) and the AxyXY-OprZ pump is associated with aminoglycoside resistance (Bador *et al.*, 2013). The whole genome sequence of strain NH44784-1996 displays several antibiotic resistance-associated genes, for example, beta-lactamase, efflux pump systems and modifying enzymes (Jakobsen *et al.*, 2013). In addition, a comprehensive analysis of resistance genes in a Chinese strain of *A. xylosoxidans* is in agreement with previous works in finding that beta-lactamases, efflux pumps and modifying enzymes are intact and resistance to the aforementioned antibiotics - penicillin, aminoglycoside and cephalosporin - is an intrinsic phenotype (Hu *et al.*, 2015). Cases of resistance to other agents in *A. xylosoxidans*, including carbapenems, have been described as acquired properties whereby the resistance genes are transferred to the pathogen via resistance genes that are carried by mobile genetic elements (Riccio *et al.*, 2001; Shin *et al.*, 2005; Neuwirth *et al.*, 2006; Di Pilato, Pollini & Rossolini, 2014; Chen *et al.*, 2014; Hu *et al.*, 2015).

5.1.5. Aims and objectives

Antibiotic resistance is considered a major problem in bacterial infections worldwide. Drug resistance patterns in *A. xylosoxidans* are interesting and challenging because few resistance phenotypes are considered to be intrinsic. Therefore, both disc diffusion and MIC will be conducted to investigate resistance in *A. xylosoxidans*.

According to Hu *et al.* (2015), several genes associated with antibiotic resistance, including beta-lactamase and efflux transport systems, were found in the *A. xylosoxidans* genomes. The whole genome sequences available for all strains used in this study will be used to predict putative genes that are associated with drug resistance. Statistical prediction will also be used to search for associations between an antibiotic resistance phenotype and related genes.

5.2. Material and Methods

5.2.1. Antibiotic susceptibility test

Antibiotic susceptibility and resistance was tested in 25 *A. xylosoxidans* isolates following guidelines for disc diffusion produced by the British Society for Antimicrobial Chemotherapy (BSAC). Antimicrobial discs were purchased from Thermo Scientific Inc. and the MAST Group. A list of the antibiotics used in this study is shown in Table 5.1. Single isolates were grown on LB agar and resuspended in sterile distilled water until the turbidity of the suspensions reached 0.5 McFarland. Each bacterial suspension was spread radially on the surface of Iso-Sensitest agar. Four to six discs were applied to each agar plate using a disc dispenser. All plates were incubated at 37°C overnight. Resulting inhibition zones were measured using a caliper. The interpretation of the zone was conducted follow BSAC standards for *Enterobacteriaceae*.

Further investigation using minimum inhibitory concentration (MIC) was carried out on Thai strains only due to their highly resistant phenotype. MIC determination was performed on a standard MIC plate, which was purchased from TREK Sensititre (Sigma Scientific), at the Clinical Microbiology Unit, Department of Pathology, Ramathibodi Hospital, Bangkok, Thailand, by courtesy of Dr. Pitak Santanirand. Mid-log phase cultures were used to inoculate into antibiotic-containing plate, each of which was labeled with the name and concentration of the antibiotic. After an overnight incubation at 37°C, the resulting MICs end-points were determined by inspection.

Table 5.1: A list of antimicrobial agents tested to determine the susceptibility of *A. xylosoxidans*. The agents are separated into chemical groups. Modes of action and known resistance mechanisms are shown.

Group	Generic name	3-letter abbreviation	Mode of action	Possible resistance mechanism
Beta-lactam				
Aminopenicillin	Amplicillin	AMP	Inhibit cell wall biosynthesis by binding transpeptidase or penicillin-binding proteins	- Enzyme modification - Efflux pump - Change in cell membrane permeability - Mutation in drug target
	Amoxicillin/Clavulanic acid	AMC		
Ureidopenicillin	Piperacillin/Tazobactam	TZP		
Cephalosporin	Cefuroxime	CXM		
	Cefotaxime	CTX		
	Ceftazidime	CAZ		
	Ceftriaxone	CRO		
	Cefepime	FEP		
Carbapenem	Imipenem	IPM		
	Meropenem	MEM		
	Ertapenem	ETP		

Table 5.1: A list of antimicrobial agents tested to determine the susceptibility of *A. xylosoxidans*. The agents are separated into chemical groups. Modes of action and known resistance mechanisms are shown. (Continued)

Group	Generic name	3-letter abbreviation	Mode of action	Possible resistance mechanism
Aminoglycoside	Gentamicin Amikacin	GEN AMK	Inhibit protein synthesis	- Enzyme modification - Efflux pump
Glycylcycline	Tigecycline	TGC	Inhibit protein synthesis	- Efflux pump
Folate biosysthesis inhibitor	Sulfamethoxazole/ Trimethoprim	SXT	Inhibit DNA replication through tetrahydrofolic synthesis	- Mutation in drug target
Fluoroquinolone	Ciprofloxacin Levofloxacin	CIP LEV	Inhibit DNA replication by binding with Topoisomerase	- Mutation in drug target - Efflux pump

5.2.2. The determination of Beta-lactamase activity

Beta-lactamase activity was tested using a common method - a nitrocefin assay. Nitrocefin is an antibiotic in the cephalosporin group, which is a sub-group of the beta-lactam antibiotics. Owing to the chromogenic properties of the drug, a normal form and a hydrolysed form have different visible colours and different UV absorbance properties. Nitrocefin is yellow and its wavelength of maximum absorbance is 390 nm, whilst the hydrolysed form is red and has a wavelength of maximum absorbance at 486 nm.

The isolates were passed through antibiotic-free medium twice to ensure that beta-lactamases were at a basal level prior to being studied. Nitrocefin (Merck-Millipore, Germany) was added at a final concentration of 51.6 µg/ml to mid-log grown cultures. Optimal density was measured at 390 nm and 486 nm every minute, for 25 minutes, using FLUOstar Omega (BMG biotech, Germany). The ratio OD_{486}/OD_{390} was calculated for every time-point to determine the activity of the beta-lactamase over the time course. Three biological replicates were performed for each isolate.

5.2.3. The use of an unsupervised learning machine to identify genes associated with a resistance phenotype

Previous studies have highlighted that drug resistance in *A. xylosoxidans* is due to both intrinsic genes and acquired genes. The search for resistance genes in *A. xylosoxidans* was performed using BLAST-based methods. The genomes of 25 isolates, newly sequenced for this study, were used in resistance genes analysis. Firstly, all genes, which were called and were annotated using Prokka version 1.7.2, were put through a reciprocal protein BLAST search against the Comprehensive Antibiotic Resistance Database (CARD) (McArthur *et al.*, 2013) with a threshold of 1e-5. Secondly, top hits from the reciprocal BLAST were validated by BLAST searches against the conserved domain database (CDD) and were manually curated to retrieve consistent results. Then, the resulting genes were assigned to previously performed OrthoMCL's output.

A matrix showing the occurrence of resistance genes, combined with drug susceptibility, was subjected to classification using the unsupervised learning algorithm, Random Forest™. Initially, Random Forest™ classification was performed using a web-based service called PhenoLink (Bayjanov *et al.*, 2012). The limitation of PhenoLink is the smallest number of genomes that contain the studied variable is 'three'. The analysis was, therefore, conducted on a stand-alone computer. The Random Forest™ package in R software was used as an alternative. The 'randomforest' package in R required two sets of data to be subjected to the analysis. The former was a dataset of resistance-associated genes only, which consisted of 64 groups from ORTHOMCL, and the latter was a whole genome dataset, which consisted of 8,762 of ORTHOMCL's groups and was used in order to validate the performance of the algorithm. The implementation followed the instruction described in Chapter 4

5.2.4. Whole genome sequencing using SMRT technology

To investigate the localisation of resistance genes-carrying integron, the genome of *A. xylosoxidans* R4 and R8 was re-sequenced using single molecule real time (SMRT) sequencing on Pacific Bioscience RS platform at CGR, and assembled sequenced reads into complete genome sequence using the HGAP software (Chin C-S, 2013). The annotation of the genomes was performed using Prokka version 1.7.2 (Seemann, 2014).

5.3. Results

5.3.1. Susceptibility phenotype

Resistance to antibiotics was preliminarily determined by disc diffusion following decision criteria provided by the BSAC. According to previous studies (Bador *et al.*, 2011, 2013), resistance to cephalosporins and aminoglycosides is intrinsic and is regulated by efflux transport systems. As summarised in Table 5.2, all isolates were less susceptible to cephalosporins, including cefuroxime (CXM), cefotaxime (CTX) and ceftriaxone (CRO); however, all isolates except R4 and R8 were still susceptible to ceftazidime (CAZ) and cefipime (FEP). Also, aminoglycosides, including gentamicin (GEN) and amikacin (AMK) were unable to inhibit the growth of the isolates (Table 5.2). Consequently, these confirmed innate resistance of *A. xylosoxidans* to certain groups of antibiotics.

When tested with aminopenicillin antibiotics, including amplicillin (AMP) and co-amoxiclav (AMC), and ureidopenicillins, including piperacillin/tazobactam (TZP), isolates from Liverpool consistently sensitive to the agents, whereas isolates from Thailand were less responsive. Strains R4 and R8 expressed resistance to amoxicillin, even with the addition of clavulanic acid, a beta-lactamase inhibitor. Considering the type strains, as expected, soil strain NCIMB 11015 responded to amoxicillin and co-amoxiclav, but clinical strain DSM 2402 did not. All isolates, except R8, were responsive to the action of piperacillin/tazobactam (Table 5.2).

Carbapenem, which is one of the most effective beta-lactam sub-groups against many Gram-negative pathogens, acted efficiently against *A. xylosoxidans* in this study (Table 5.2). Nevertheless, two Thai clinical isolates, R4 and R8, displayed high resistance to all of the carbapenem antibiotics used in this study, including imipenem (IPM), meropenem (MEM) and ertapenem (ETP).

Tigecycline (TGC), a glycylcycline that was launched to the market in 2005, is a new generation of tetracycline, which acts by preventing amino acyl tRNA from

sitting on ribosomes. As shown in Table 5.2, tigecycline was the only drug that is still active against *A. xylosoxidans* isolates, although strains DSM 2402, R7 and R8 were less sensitive to the drug.

Most clinical isolates from Liverpool and lab strains demonstrated susceptibility to fluoroquinolones, but isolates from Thailand and a clinical type strain from Japan were less susceptible to drugs in this category. Of the Liverpool strains (Table 5.2), L15 and L16 were less sensitive to ciprofloxacin (CIP), a second-generation agent, but they were still responsive to the activity of levofloxacin (LEV), a third-generation agent. Strains from East Asia were less susceptible to CIP, whereas they were still susceptible to LEV. Strains DSM 2402 and R8 were the only strains that were resistant to LEV. Strain R6 was only the strain that was susceptible to the fluoroquinolones (Table 5.2). This evidence suggested that *A. xylosoxidans* in East Asia, Thailand in particular is resistant to fluoroquinolones.

The last group of antibiotics tested in this study was the combination of folic acid biosynthesis inhibitors, sulfamethoxazole and trimethoprim (SXT). Both Thai and Liverpool isolates were susceptible to the action of SXT (Table 5.2). With consistent susceptibility, strains R4 and R8 did not respond to the bactericidal activity of the agents. The environmental strain NCIMB 11015 was susceptible to the agents, yet the clinical strain DSM 2402 was not.

These preliminary results show that clinical isolates from Thailand demonstrated greater antibiotic resistance than clinical isolates from Liverpool did. Two Thai isolates, R4 and R8, were multidrug resistant strains because these two isolates showed resistance to almost all of tested antibiotics. This is in agreement with previous studies (Bador *et al.*, 2011, 2013; Hu *et al.*, 2015), which found that the resistance to cephalosporin and aminoglycoside is an intrinsic property. TZP and SXT, common clinically-used antimicrobial agents, were still active against *A. xylosoxidans* strains, except R4 and R8. Due to the limitations of disc diffusion, which could not illustrate the dosage of each antibiotic that inhibits the growth of bacteria, further investigation into the inhibitory doses of antibiotics was conducted on the group of Thai clinical isolates.

The determination of MICs was carried out for isolates from Thailand only, in order to quantify levels of resistance. As shown in Table 5.3, susceptibility of Thai isolates was consistent with results from the disc diffusion method in which the strains were less susceptible to cephalosporins and aminoglycosides. CAZ was the only cephalosporin used in this study that was able to inhibit the growth of Thai strains of *A. xylosoxidans.* The aminoglycosides were less effective against *A. xylosoxidans*. This was in agreement with previous studies (Bador *et al.*, 2011, 2013), reaffirming the resistance of *A. xylosoxidans* to cephalosporins and aminoglycosides.

Whilst MICs for aminopenicillins and carbapenems fell into the susceptible category in most Thai isolates, strain R4 and strain R8 showed resistance to those antimicrobial agents (Table 5.3). These two strains were suspected of expressing an extended-spectrum beta-lactamase that is able to hydrolyse carbapenem antibiotics. Therefore, further investigation into the presence of the beta-lactamase enzyme in Thai isolates should be conducted to reveal the underlying mechanism of extended-spectrum beta-lactam resistance.

The determination of MICs for fluoroquinolones gave a result that is consistent with the disc diffusion method, suggesting that Thai isolates developed resistance to CIP but are still susceptible to LEV (Table 5.2 and Table 5.3). Regarding susceptibility to LEV, R1, R4 and R5 remained responsive to the drug whilst the others were less susceptible and resistant to the drug (Table 5.3). Of the Levofloxacin-resistant group, strain R8 had the highest MIC (Appendix Table A3.4), which meant that R8 would probably have a unique mechanism, which will be discussed in the section entitled ‘Other resistance-associated proteins’.

Table 5.2: Resistance in *A. xylosoxidans* to commonly-used antibiotics, tested using disc diffusion. The table presents an interpretation of susceptibility, determined using cut-off levels based on BSAC guidelines. The susceptibility interpretation is illustrated as 'Resistant (R, Red)', 'Intermediate (I, Yellow)' and 'Susceptible (S, Green)'. NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11014; DSM 2402 for *A. xylosoxidans* DSM 2402, and L-strains are Liverpool's clinical isolates.

Isolates	AMP	AMC	TZP	CXM	CTX	CAZ	CRO	FEP	IPM	MEM	ETP	LEV	CIP	GEN	AMK	SXT	TGC
NCIMB 11015	S	S	S	R	R	S	R	I	S	S	S	S	S	R	I	S	S
DSM 2402	R	R	S	R	R	S	R	R	S	S	S	R	R	R	R	R	R
L1	S	S	S	R	R	S	R	I	S	S	S	S	S	R	I	S	S
L5	S	S	S	R	R	S	R	R	S	S	S	S	S	R	R	S	S
L8	S	S	S	R	R	S	R	R	S	S	I	S	S	I	I	S	S
L10	S	S	S	R	R	S	R	I	S	S	S	S	S	R	I	S	S
L11	S	S	S	R	R	S	S	I	S	S	S	S	S	S	S	S	S
L14	S	S	S	R	R	S	R	R	S	S	S	S	S	R	S	S	S
L15	S	S	S	R	R	S	I	S	S	S	S	S	R	R	S	S	S
L16	S	S	S	R	R	S	R	R	S	S	S	I	R	R	R	S	S
L17	S	S	S	R	R	S	R	R	S	S	S	S	S	S	S	S	S

Table 5.2: Resistance in *A. xylosoxidans* to commonly-used antibiotics, tested using disc diffusion. The table presents an interpretation of susceptibility, determined using cut-off levels based on BSAC guidelines. The susceptibility interpretation is illustrated as 'Resistant (R, Red)', 'Intermediate (I, Yellow)' and 'Susceptible (S, Green)'. R-strains are Thai clinical isolates. (Continued)

Isolates	AMP	AMC	TZP	CXM	CTX	CAZ	CRO	FEP	IPM	MEM	ETP	LEV	CIP	GEN	AMK	SXT	TGC
R1	I	I	S	R	R	S	R	I	S	S	S	S	I	R	R	S	S
R2	S	S	S	R	R	S	R	S	S	S	S	S	I	R	R	S	S
R3	I	S	S	R	R	S	R	I	S	S	S	S	I	R	R	S	S
R4	R	R	S	R	R	R	R	R	I	R	R	S	I	R	R	R	S
R5	I	I	S	R	R	S	R	I	S	S	S	S	I	R	R	S	S
R6	S	S	S	R	R	S	S	R	S	S	S	S	S	R	S	S	S
R7	S	S	S	R	R	S	I	R	S	S	S	S	R	R	R	S	I
R8	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	I
R9	S	S	S	R	I	S	I	S	S	S	S	S	I	R	I	S	S
R10	I	S	S	R	R	S	R	I	S	S	S	S	R	R	R	S	S
R11	S	S	S	R	R	S	R	R	S	S	S	S	S	R	S	S	S
R12	S	S	S	R	R	S	R	S	S	S	S	S	I	R	I	S	S
R13	I	S	S	R	R	S	R	I	I	S	S	I	R	R	R	S	S
R14	I	S	S	R	R	S	R	I	S	S	S	S	I	R	R	S	S

Table 5.3: Interpretation of susceptibility in Thai isolates of *A. xylosoxidans* to antimicrobial agents based upon determination of the minimum inhibitory concentration (MIC). The susceptibility interpretation is illustrated as 'Resistant (R, Red)', 'Intermediate (I, Yellow)' and 'Susceptible (S, Green)'. R-strains are Thai clinical isolates.

Isolates	AMP	AMC	TZP	CXM	CTX	CAZ	CRO	FEP	IPM	MEM	ETP	CIP	LEV	GEN	AMK	SXT	TGC
R1	R	S	S	R	R	I	R	R	I	S	S	R	S	R	R	S	S
R2	S	S	S	R	R	I	R	R	S	S	S	R	R	R	R	S	S
R3	S	S	S	R	R	I	R	R	S	S	S	R	I	R	R	S	S
R4	R	R	S	R	R	R	R	R	I	R	R	I	S	R	R	R	S
R5	R	S	S	R	R	I	R	R	I	S	S	R	S	R	R	S	S
R6	S	S	S	R	R	S	R	I	S	S	S	R	I	I	I	S	S
R7	S	S	S	R	R	I	R	R	S	S	S	R	R	R	R	S	S
R8	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	S
R9	S	S	S	R	R	I	R	R	S	S	S	R	I	R	R	S	S
R10	S	S	S	R	R	I	R	R	S	S	S	R	I	R	R	S	S
R11	S	S	S	R	R	I	R	R	S	S	S	R	R	R	R	S	S
R12	S	S	S	R	R	I	R	R	S	S	S	R	R	R	R	S	S
R13	R	R	S	R	R	I	R	R	S	S	S	R	R	R	R	S	S
R14	R	S	S	R	R	I	R	R	I	S	S	R	I	R	R	S	S

The carbapenem-hydrolysing beta-lactamase, or carbapenemase, is a versatile beta-lactamase that is able to hydrolyse various groups of beta-lactam antibiotics, such as the penicillins, cephalosporins and carbapenems. As such, the investigation of the effect of carbapenemase upon clinical isolates is essential for clinicians to select an antibiotic of choice. The investigation into the beta-lactamase's activity was investigated using a nitrocefin assay. After incubating cultures with nitrocefin for 25 minutes, visible colour changes of the suspensions were inspected as illustrated in Appendix Figure A3.1. The colour of nitrocefin in R4 and R8 cultures changed from yellow (native form) to red (hydrolysed form).

In most of the isolates tested, except R4, R8 and R13, the conversion of nitrocefin progressed slowly. The ratio, OD_{486}/OD_{390}, changed from approximately 0.5 to 0.6 (Figure 5.5). Strain R13 slowly transformed nitrocefin into a hydrolysed form. Its ratio slowly shifted from 0.4 to 0.9 in 25 minutes. Interestingly, 2 Thai MDR strains demonstrated a high rate of nitrocefin hydrolysis. The colour of the suspensions in R4's and R8's wells changed immediately when nitrocefin was put into the wells. The ratio of optical density of strain R4 dramatically increased in the first minute after incubation and it levelled off at approximately 1.1 until the taking of measurements stopped. The ratio of optical density of strain R8 rose at a constant rate until it reached 1.1 at the point that the taking of measurements stopped.

Using the ratios at the end of the measurement-taking, an ANOVA test revealed a significant difference between the ratios seen for the strains ($F_{2,39} = 50.5$; $p < 0.001$). A post-hoc test using Tukey HSD showed significant differences between strain R4 and the others ($p < 0.001$), and between R8 and the others ($p < 0.001$), as presented in Figure 5.6. This suggested that the performances of the beta-lactamases of R4 and R8 were more vigorous than those of the other strains.

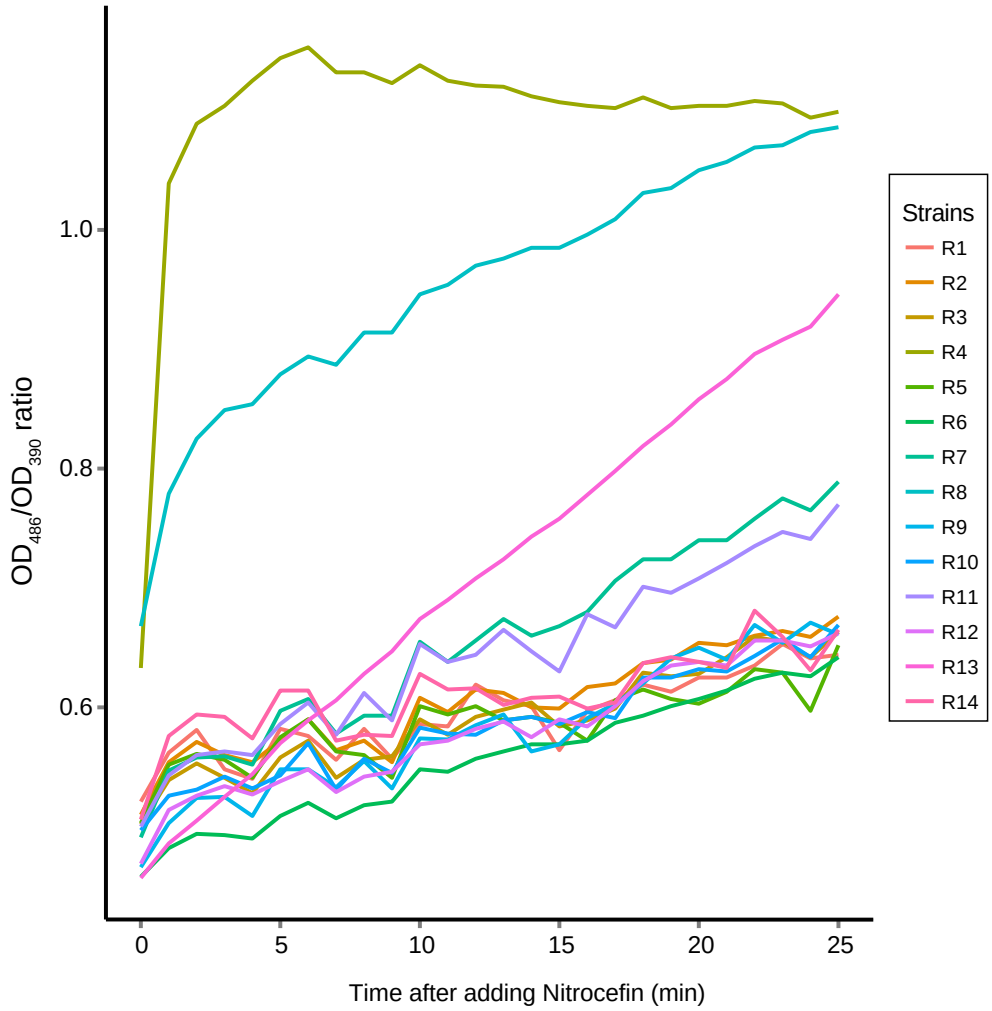


Figure 5.5: Changes in OD_{486}/OD_{390} ratios for clinical Thai isolates of *A. xylosoxidans* after incubation with nitrocefin for 25 minutes. The X-axis indicates the length of incubation after the addition of nitrocefin. The Y-axis indicates the change of OD_{486}/OD_{390} ratios for each isolate. R1-R14 nomenclature represents Thai clinical isolates 1-14.

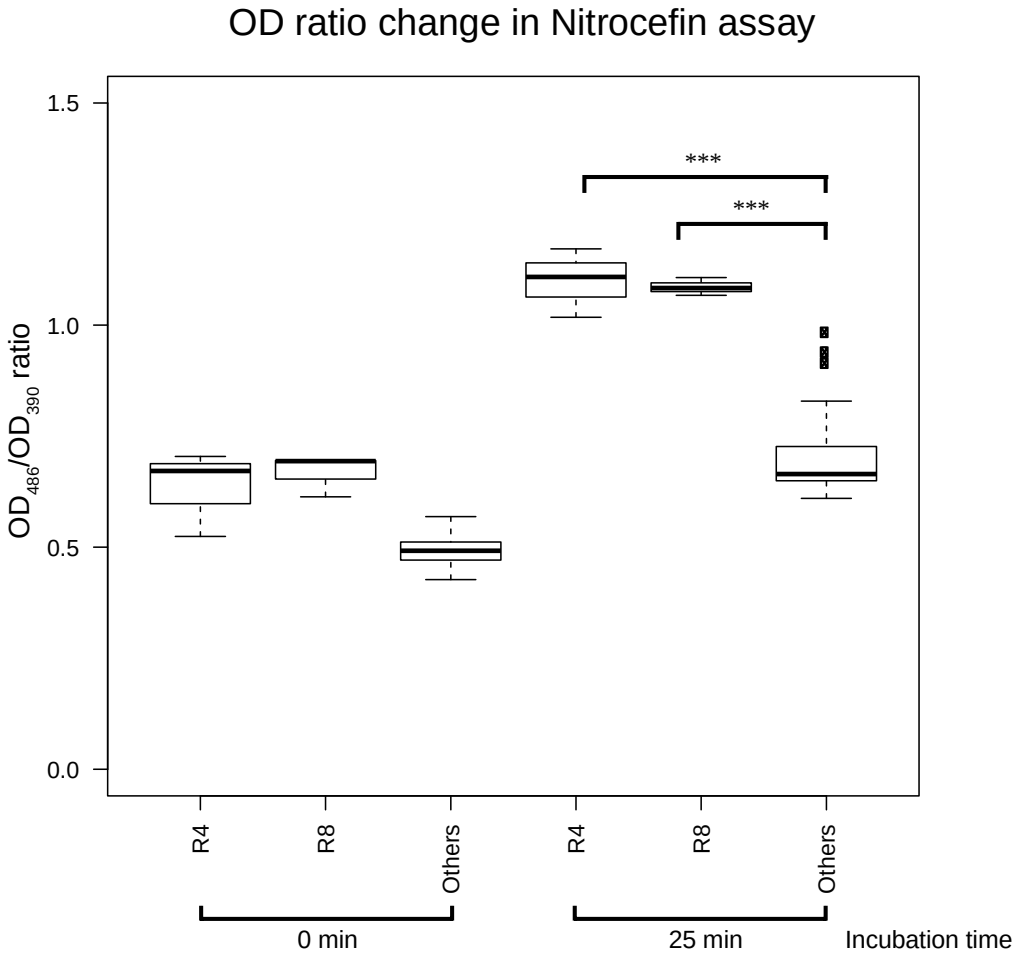


Figure 5.6: The changes in OD_{486}/OD_{390} ratios in carbapenem-resistant clinical Thai *A. xylosoxidans* strains R4 and R8, compared to susceptible strains, after incubation with Nitrocefin for 25 minutes. Optical density of each strain was obtained from 3 independent experiments. Means of the ratio changes for strain R4 and strain R8 are significantly different (Tukey's HSD, p-value < 0.01)

5.3.2. Antibiotic resistance-associated genes

Following a protein BLAST search against CARD and orthologous gene grouping, a total number of 1,303 resistance genes, taken from 26 genomes, were grouped into 64 groups, based on their amino acid sequence similarity, using OrthoMCL (Appendix Table A3.6). The largest group was ORTHOMCL0, which consisted of 154 genes from 26 genomes, and it was annotated as 'RND efflux system, inner membrane transporter CmeB'. The second largest group, ORTHOMCL1, contained 87 genes from 26 genomes and its annotation was 'AcrB/AcrD/AcrF family protein'. The third largest was ORTHOMCL2, which consisted of 30 genes from 25 genomes. These 3 groups were the only groups that contained more than one gene from one genome.

5.3.3. Phylogenetic tree of conserved antibiotic resistance genes

As shown in Table 5.4, comparative genomic analysis revealed that 31 groups of putative drug resistance genes were commonly shared across *A. xylosoxidans*; seven were beta-lactamases, 17 were efflux pumps, and seven were resistance-associated proteins (Table 5.4).

Phylogenetic relationships using a maximum-likelihood algorithm (Figure 5.7), described the clustering of strains, bearing similarities to the core genome phylogeny (Figure 4.5, Figure 4.6). To investigate the relationship between the core genome phylogeny and conserved antibiotic resistance phylogeny, a tanglegram tree (Figure A3.2) suggested that the conserved resistance phylogeny partially explains the core phylogeny. In other words, conserved antibiotic resistance gene phylogeny could not represent core genome phylogeny.

Table 5.4: Antibiotic resistance-associated genes that have been identified as conserved genes in the genomes of 26 strains of *A. xylosoxidans*

		Reference gene ID	Gene annotation
Beta-lactamase		YP_008033064.1	Beta lactamase class C and other penicillin binding proteins
		YP_008032142.1	Beta lactamase class C and other penicillin binding proteins
		YP_008032848.1	Beta lactamase AmpC
		YP_008031641.1	Beta lactamase
		YP_008028833.1	Metallo beta lactamase family protein, putative
		YP_008031922.1	Metallo beta lactamase superfamily protein
		ADO14463.1	Oxacillinase OXA-114
Efflux transport system	RND efflux	YP_008029803.1	RND efflux system, inner membrane transporter CmeB
		YP_008031292.1	RND efflux system, inner membrane transporter CmeB
		YP_008033424.1	RND efflux system, inner membrane transporter CmeB
		YP_008029804.1	RND efflux system, membrane fusion protein CmeA
		YP_008027730.1	RND efflux system, membrane fusion protein CmeA
		YP_008031291.1	RND efflux system, membrane fusion protein CmeA
		YP_008033423.1	RND efflux system, outer membrane lipoprotein CmeC
		YP_008028521.1	RND multidrug efflux transporter; Acriflavine resistance protein
		YP_008027735.1	RND multidrug efflux transporter; Acriflavine resistance protein
		YP_008030766.1	Inner membrane component of tripartite multidrug resistance system
	ABC efflux	YP_008033182.1	ABC transporter, ATP binding permease protein
		YP_008031007.1	Lipid A export ATP binding permease protein MsbA
		YP_008027505.1	Lipid A export ATP binding permease protein MsbA
		YP_008027185.1	Lipid A export ATP binding permease protein MsbA
		YP_008028300.1	Macrolide export ATP binding permease protein MacB
		YP_008031184.1	Transport ATP binding protein CydCD
	MFS	YP_008031014.1	Fosmidomycin resistance protein
	Other	YP_008033356.1	Transcriptional regulatory protein OmpR
Resistance-associated enzyme		CKH34743.1	Chloramphenicol acetyltransferase
		YP_008031014.1	Fosmidomycin resistance protein
		CKH16534.1	Topoisomerase IV subunit B
		YP_008030648.1	Dihydropteroate synthase
		AIR49201.1	DNA gyrase subunit B
		CKH08160.1	DNA gyrase subunit A
		YP_008028255.1	Aminoglycoside-N^3-acetyltransferase

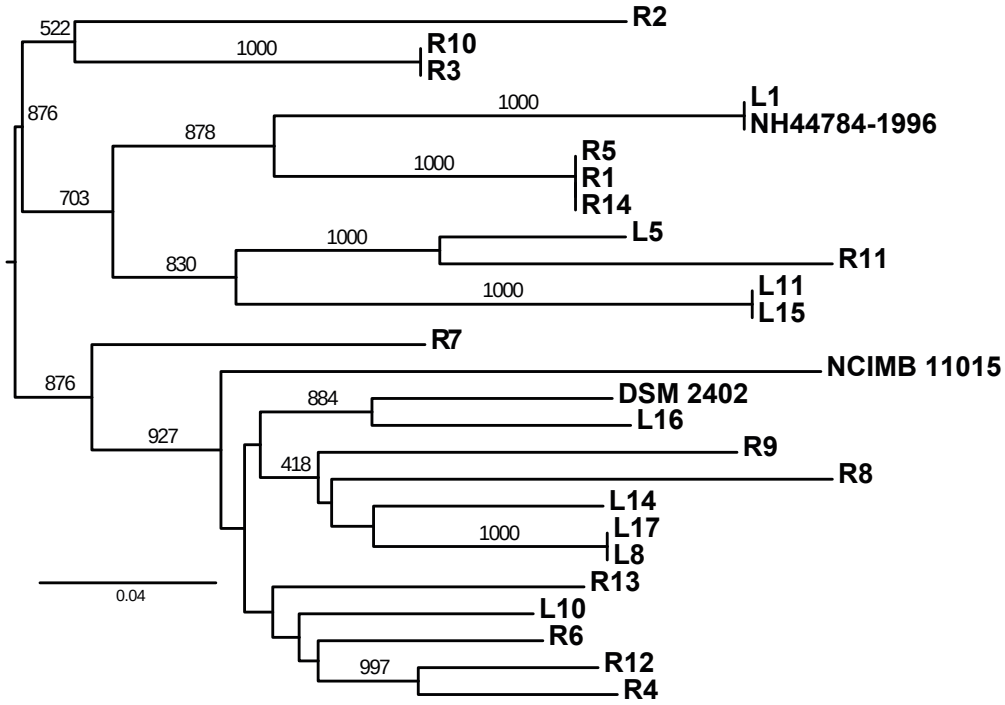


Figure 5.7: Phylogenetic trees based on antibiotic resistance-associated genes that are conserved across 26 genomes of *A. xylosoxidans.* Maximum-likelihood tree constructed with 1,000 bootstrap repeats. Bootstrap values more than 50% were indicated. NH44784-1996 stands for *A. xylosoxidans* NH44784-1996; NCIMB 11015 for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402. L-strains are Liverpool's clinical isolates and R-strains are Thai clinical isolates.

5.3.4. Efflux pumps

A group of efflux transport systems in *A. xylosoxidans* was identified as the largest group of resistance-associated genes, including 18 intrinsic genes (Table 5.4) and 28 acquired genes (Table 5.5 and Appendix Table A3.7). Amongst the eighteen pump-associated intrinsic genes identified, ten were associated with resistance nodulation cell division (RND) pumps, six were ATP-binding cassette (ABC) pumps, one was a major facilitator superfamily (MFS) pump and one was a porin regulator OmpR (Table 5.5).

The first group of known antibiotic resistance-associated RND efflux system was CmeABC, which is consisted of CmeA (Reference gene id: YP_008027730.1), a membrane fusion protein, CmeB (Reference gene id: YP_008027731.1), an inner membrane transporter, and CmeC (Reference gene id: YP_008027732.1), an outer membrane protein. Only CmeA was found across all 26 strains (Table 5.4). The CmeB subunit and CmeC subunit were present in 23 strains and 25 strains, respectively (Appendix Table A3.8). The absence of the CmeB and CmeC genes in some strains was probably due to an error during the assembly or sequencing processes.

This CmeC efflux pump system shared 96.9% similarity with the AxyXY-OprZ RND-type pump in *A. xylosoxidans* AXX-A (renamed *A. insuavis* AXX-A) (Appendix Figure A3.3). Bador *et al.* (2013) described the AxyXY-OprZ operon as being involved in resistance to carbapenems, aminoglycosides and fluoroquinolones.

Table 5.5: Antibiotic resistance-associated genes that have been identified as acquired genes in the genomes of 26 strains of *A. xylosoxidans*

		Reference gene ID	Gene annotation
Beta-lactamase		AHY18660.1	IMP-14 metallo beta lactamase
		YP_008027752.1	Beta lactamase
		YP_008029689.1	Beta lactamase
		AIR52250.1	Beta lactamase domain containing protein
		CDI94943.1	Beta lactamase domain containing protein
		YP_008031839.1	Metallo beta lactamase family protein, RNA specific
Efflux transport system	RND	WP_020925327.1	AcrB/AcrD/AcrF family protein
		YP_008028715.1	AcrB/AcrD/AcrF family protein
		WP_013393781.1	Acriflavine resistance protein B
		YP_008027731.1	RND efflux system, inner membrane transporter CmeB
		YP_008033425.1	RND efflux system, membrane fusion protein CmeA
		YP_008028035.1	RND efflux system, membrane fusion protein CmeA
		YP_008027732.1	RND efflux system, outer membrane lipoprotein CmeC
		YP_008032982.1	RND efflux system, outer membrane lipoprotein CmeC
		YP_008027734.1	RND efflux system, outer membrane lipoprotein CmeC
		YP_008029977.1	RND efflux system, outer membrane lipoprotein CmeC
		YP_008031293.1	RND efflux system, outer membrane lipoprotein CmeC
		YP_008032981.1	RND multidrug efflux transporter; Acriflavine resistance protein
		WP_020925545.1	Cobalt-zinc-cadmium resistance protein CzcA
		CKG72640.1	Cation efflux system protein CzcA
		YP_008032003.1	RND multidrug efflux transporter; Acriflavine resistance protein
		YP_008033107.1	Inner membrane component of tripartite multidrug resistance system
		YP_008030767.1	Membrane fusion component of tripartite multidrug resistance system
	ABC	CKG97466.1	Iron import ATP-binding/permease protein IrtB
		YP_008030221.1	Putative ABC iron siderophore transporter, fused permease and ATPase domains
		CKH86978.1	Putative multidrug export ATP-binding/permease protein SAV1866
		YP_008032049.1	Cyclolysin secretion ATP binding protein
		YP_008031197.1	Transport ATP binding protein CydCD
	MFS	YP_008028726.1	MFS permease
		AKE04698.1	MFS Transporter
		CKI18085.1	Spectinomycin/tetracycline efflux pump
		YP_008031167.1	Tetracycline efflux protein TetA
	Other	WP_019396923.1	Cation transporter
		ADP16465.1	Periplasmic linker protein 2

Table 5.5: Antibiotic resistance-associated genes that have been identified as acquired genes in the genomes of 26 strains of *A. xylosoxidans* (Continued)

	Reference gene ID	Gene annotation
Resistance-associated enzyme	YP_008029342.1	Translation elongation factor Tu
	YP_008029360.1	Translation elongation factor G
	YP_008029349.1	DNA directed RNA polymerase beta subunit
	YP_008030789.1	Topoisomerase IV subunit A
	YP_008031164.1	Putative streptomycin phosphotransferase
	AHY18660.1	Dihydropteroate synthase
	BAN89135.1	Streptomycin-6-kinase
	AIA53578.1	3'-kinase
	AHY18660.1	Aminoglycoside-N^6-acetyltransferase

A breakdown analysis of this AxyXY-OprZ-homologous CmeABC efflux pump revealed the cluster of pumps that is associated with this level of drug resistance. Determining that the CmeB subunit is a functional subunit of the pump, a phylogenetic tree of CmeB genes showed a cluster that consists of CmeB genes from strains L11, L15, and R11 (Figure 5.8), which are strains that exhibit low levels of aminoglycoside resistance (Appendix Table A3.2). Moreover, analysis of the pan-genome of antibiotic resistance in genus *Achromobacter* emphasised the association between this pump and aminoglycoside resistance in *Achromobacter* as the genes encoding CmeABC efflux were not present in *A. piechaudii* and *A. spanius.* Interestingly, the CmeABC efflux pump genes were present in the aminoglycoside-susceptible *Achromobacter* species, *A. insolitus* and *A. ruhlandii*, but they were clustered distantly from the genes from aminoglycoside-resistant *Achromobacter* species (Figure 5.9). Phylogenetic relationships revealed that the aminoglycoside resistance in genus *Achromobacter* has developed from a susceptible phenotype to resistant phenotype.

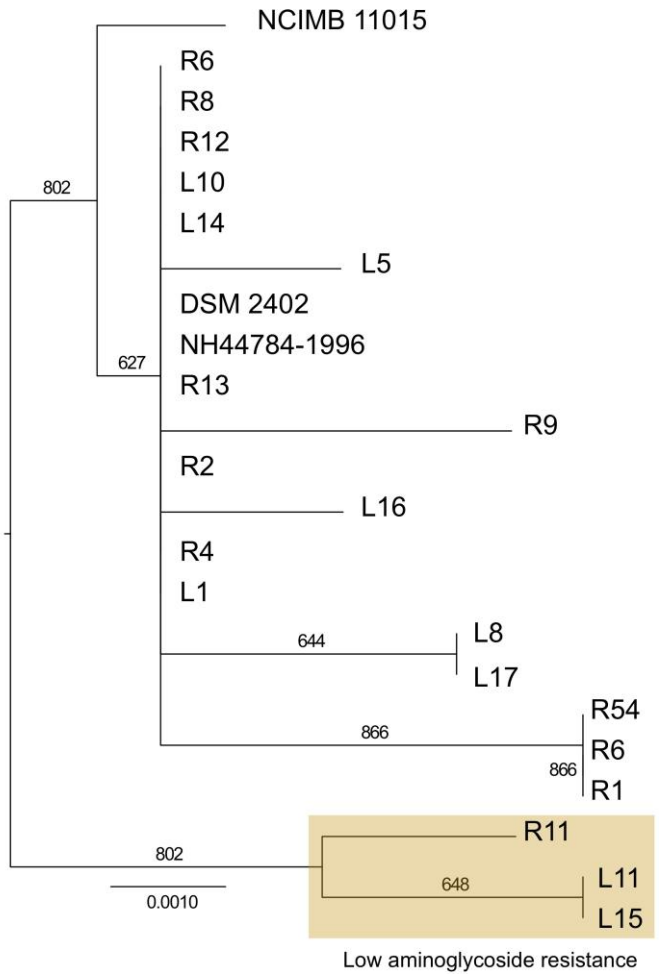


Figure 5.8: A Neighbour-Joining phylogenetic tree based on aminoglycoside-resistance-associated CmeB in *A. xylosoxidans*. The tree was constructed with 1,000 bootstrap repeats. The tree is constructed NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; NH44784-1996 for *A. xylosoxidans* NH44784-1996; R-strains are Thai clinical isolates; and L-strains are Liverpool's clinical isolates. Bootstrap values more than 75% were indicated.

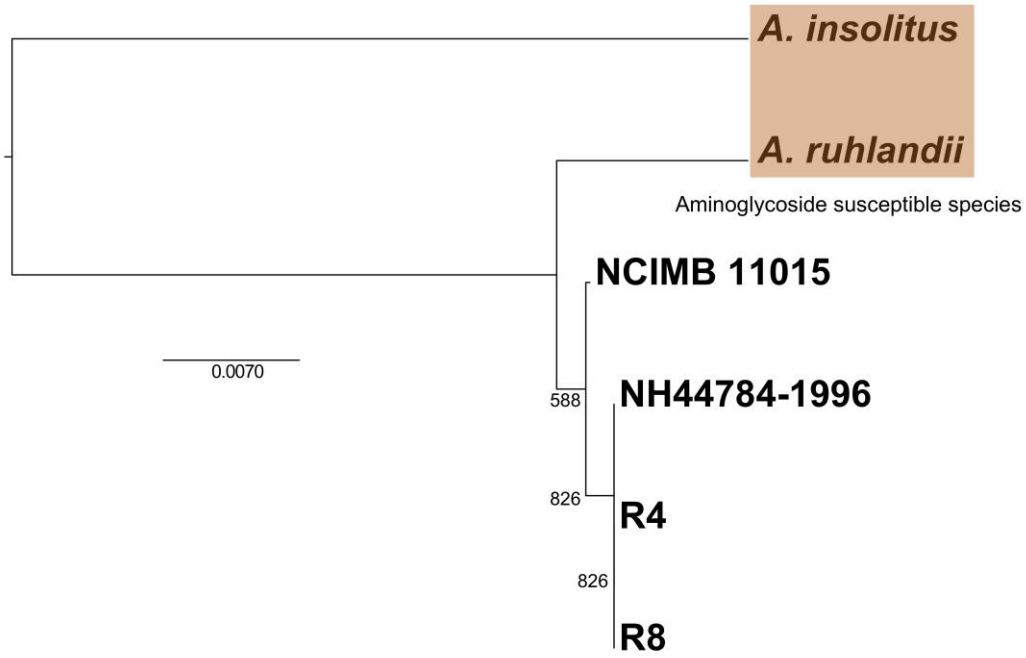


Figure 5.9: A Neighbour-Joining phylogenetic tree based on aminoglycoside-resistance-associated CmeB in genus *Achromobacter*. NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; NH44784-1996 for *A. xylosoxidans* NH44784-1996; and R-strains are Thai clinical isolates. Bootstrap values > 50% were indicated.

In addition, another paralogue of these known efflux transport systems, CmeABC, was made of CmeA (Reference gene id: YP_008031291.1), CmeB (Reference gene id: YP_008031292.1) and CmeC (Reference gene id: YP_008031293.1). Both CmeA and CmeB were conserved across 26 genomes in this study (Appendix Table A3.9). This CmeABC shared 98.4% homology with the AxyABM efflux pump in *A. insuavis* AXX-A (Bador *et al.*, 2011)(Appendix Figure A3.4). This pump plays an important role in cephalosporin and fluoroquinolone resistance. The presence of this efflux system in *A. xylosoxidans* was in agreement with the antibiotic resistance profiles of the isolates, as illustrated in Table 5.2 and 5.3. Likewise, this efflux pump was associated with cephalosporin resistance in other *Achromobacter* species.

The group containing an acriflavine resistance protein was another group of RND efflux pumps found in *A. xylosoxidans* (Table 5.4 and 5.5). Two copies of acriflavine resistance proteins were expressed in the core genome of *A. xylosoxidans*. Acriflavine resistance associated genes were also identified in accessory resistance genes (Table 5.5). In *E. coli*, Acriflavine resistance protein B (AcrB) is an essential factor for the survival of the bacteria. AcrB gathers with 2 other proteins, AcrA and TolC (multifunctional channel), to form a complex efflux system, AcrAB-TolC. This efflux pump plays an important role in protecting *E.coli* from organic solvents such as nonane (Tsukagoshi & Aono, 2000). In the comparative genomic analysis of this study, AcrB and TolC were present in the genome but the relationship between them was not observed in strains used in this study as well as in strain NH44784-1996 (Jakobsen *et al.*, 2013).

Eleven ATP-binding cassette (ABC) efflux pumps were identified across 26 genomes of *A. xylosoxidans*: six intrinsic genes (Table 5.4) and five acquired genes (Table 5.5). The majority of the intrinsic ABC pumps were involved in macrolide resistance. Two macrolide-resistance groups were intrinsically present in *A. xylosoxidans*: the Lipid A exporter and the macrolide exporter MacB. The main function of the MsbA protein is to export Lipid A, which is a part of the lipopolysaccharide (LPS), to outside of the cells. The interaction between this pump and some antibiotics, such as macrolides, has resulted in drug resistance (Woebking *et al.*, 2005). Interestingly, MsbA has two binding sites, which can interact with Lipid A and drugs separately (Siarheyeva & Sharom, 2009). The macrolide-specific exporter is a transmembrane protein that extrudes macrolide antibiotics from cells, in cooperation with MacB and TolC (Kobayashi, Nishino & Yamaguchi, 2001).

The only MFS–containing efflux pump presented as intrinsic was the fosmidomycin resistance protein. A study by Fujisaki *et al.* (1996) revealed that this protein has an effect on fosmidomycin resistance in *E. coli*. Fosmidomycin is an inhibitor of 2-C-methyl-D-erythritol-4-phosphate (MEP), which is an enzyme in the first step of the non-mevalonate pathway of isoprenoid biosynthesis. This

pathway is important in some human pathogens, such as *A. baumannii*, *P. aeruginosa* (Eberl *et al.*, 2003) and *A. xylosoxidans* (from this analysis). The presence of the fosmidomycin resistance protein in *A. xylosoxidans* suggests that the bacteria have the potential to be fosmidomycin-resistant.

In total, four groups of efflux pumps containing MFS were expressed, as acquired genes, by *A. xylosoxidans*. Two tetracycline-associated efflux pumps - spectinomycin/tetracycline efflux and Tetracycline efflux protein tetA - were identified; however, spectinomycin and tetracycline were not used in this study. The presence of these MFS efflux pumps was not related to tigecylcine resistance.

5.3.5. Beta-lactamase

Since it has been reported that *A. xylosoxidans* is resistant to multiple antibiotics, especially antibiotics in the beta-lactam group, genes associated with beta-lactam resistance were identified using CARD and validated using the NCBI database. A total of 13 beta-lactamase orthologous groups were revealed by OrthoMCL. Seven out of these thirteen groups were conserved across 26 *A. xylosoxidans* genomes (Table 5.4).

These conserved beta-lactamase genes consisted of three groups of class C beta-lactamase (one AmpC beta-lactamase and two general class C beta-lactamases), two groups of metallo-beta-lactamase, one group of class D beta-lactamase, and one group of non-specified beta-lactamase (Table 5.4). Only the class D beta-lactamase, or oxacillinase-114, group has been well identified and biochemically characterised in *A. xylosoxidans* (Turton *et al.*, 2011). Finding these intrinsic beta-lactamase genes ascertained that resistance to beta-lactam antibiotics is intrinsic in *A. xylosoxidans*.

Acquired beta-lactamases were composed of six beta-lactamase genes. There were two beta-lactamase domain-containing proteins, one metallo-beta-lactamase, one class A beta-lactamase, one unclassified beta-lactamase and one RNA-

specific metallo-beta-lactamase. The class A beta-lactamase is usually present in Gram-negative bacteria and is a target for many beta-lactamase inhibitors, including clavulanic acid and tazobactam.

Resistance to aminopenicillin antibiotics was observed in these 25 strains of *A. xylosoxidans*. Of 74 annotated resistance-associated genes in the species, Random Forest™ ranked a beta-lactamase gene (reference gene id: YP_008027752.1) first with the highest values in both mean decreased accuracy score and mean decreased Gini score (Figure 5.10). This protein was annotated as a Class A beta-lactamase. Most beta-lactamase inhibitors target class A beta-lactamases. This gene was found in strains that less susceptible to aminopenicillin: L1, R1, R3, and R5. With reference to Table 5.2 and Table 5.3, the addition of clavulanic acid made aminopenicillin work more effectively than aminopenicillin alone. This indicated an association between this selected beta-lactamase and aminopenicillin resistance.

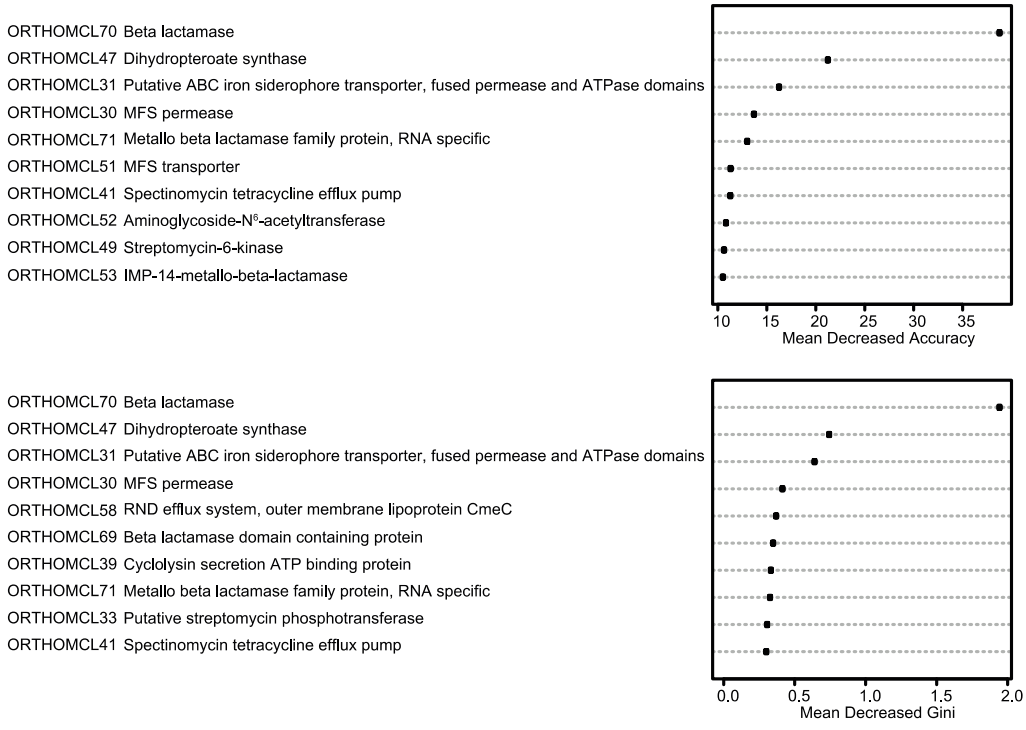


Figure 5.10: A variable importance plot based on Random Forest™ classification of AMP susceptibility in 25 isolates of *A. xylosoxidans*, showing the importance of each orthologue. The x-axes indicate the level of importance. The Y-axes show the list of orthologues.

An interesting finding was that the metallo-beta-lactamase found only in R4 and R8 shared 100% similarity to an IMP-14 metallo-beta-lactamase found in an *A. baumannii* isolate from the same hospital (Appendix Figure A3.5). Strains R4 and R8 were the only strains that were resistant to carbapenems (Table 5.2 and Table 5.3). This suggested that the finding of IMP-14 metallo-beta-lactamase would be associated with the resistance to carbapenems in strains R4 and R8.

To investigate the presence of a carbapenemase gene in Thai MDR *A. xylosoxidans* genomes, found to be unique when the genomes were compared to the reference genome, reads from the MDR strains that did not map to the reference were assembled. The assembled reads that did not map to the reference genome showed that an IMP-14-metallo-beta-lactamase gene was present in the genomes of R4 and R8, whereas it was not present on the reference genome or the genomes of the other strains. When nucleotide sequences were viewed, the gene was found to be surrounded by other genes, such as integrase, and conserved sequences, such as attI and attC, that constitute an integron. In addition, this structure was detected in the whole genome assemblies of R4 and R8.

With Pacific Biosciences' single molecular real-time sequencing technology, the resistance gene-carrying integron was found in the genomes of R4 and R8. It consisted of antibiotic resistance genes and conserved sequences (Genbank accession no. KJ406505-KJ406506) (Figure 5.11). This integron was only found in strains R4 and R8, not featuring in other Thai MDR strains (Figure 5.11).

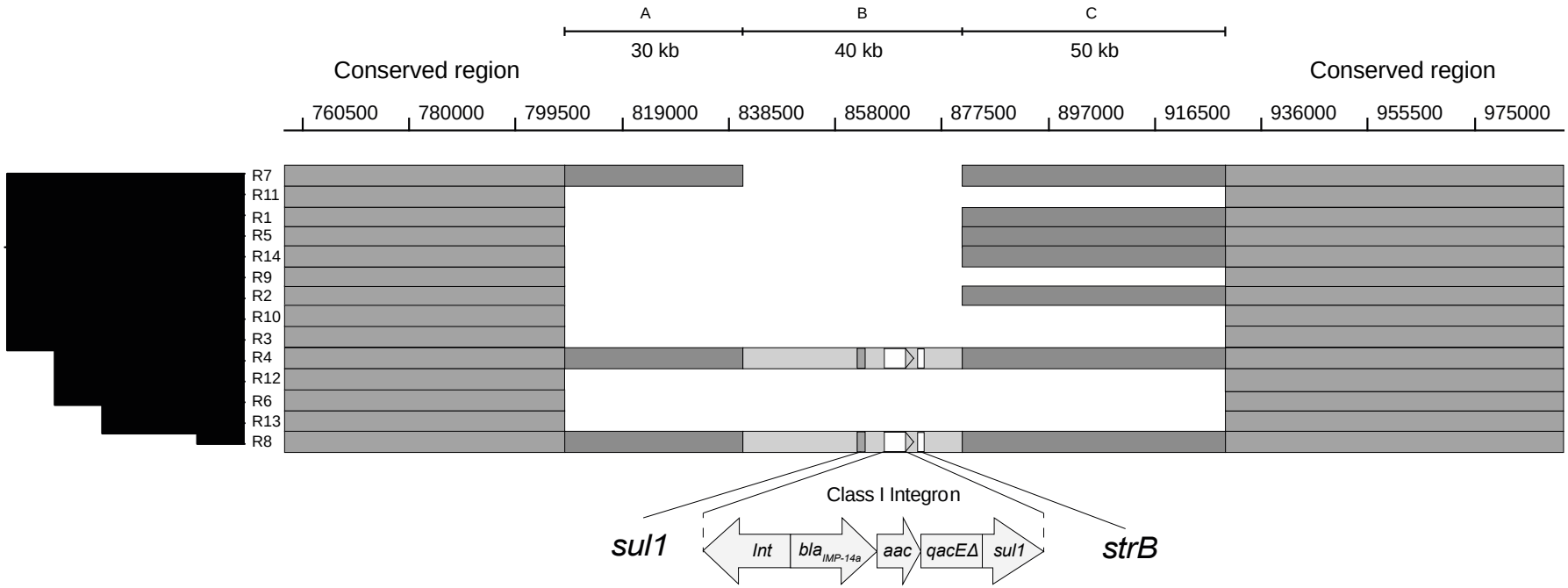


Figure 5.11: The insertion of a 40kb-long class 1 integron into the chromosome of Thai *A. xylosoxidans* strains R4 and R8. The class 1 integron consists of conserved genes for integron structure: *Int*, *qacΔE* and *sul1*. Genes in the cassette are $bla_{OXA-114}$ and *aac(6')*. The dendrogram shows the phylogenetic relationships of Thai *A. xylosoxidans* isolates based on a MLST scheme. R1-R14 are Thai clinical isolates. *Int* stands for Integrase; $bla_{IMP-14a}$ for IMP-14a metallo-beta-lactamase; *aac* for aminoglycoside acetyltransferase; *qacEΔ* for quaternary ammonium compound transporter; and *sul1* for sulfonamide-resistance dihydropteroate synthase.

The gene cassette in the integron was composed of the following genes: *int* - $bla_{OXA-114}$ - *aac(6') – qacΔE – sul1*. This kind of structure has been previously reported in Thailand (Piyakul, Tiyawisutsri & Boonbumrung, 2012). A comparative study revealed that the nucleotide sequences of the integrons found in R4 and R8 shared 100% similarity (Appendix Figure A3.5). Interestingly, the alignment of the $Bla_{OXA-114}$ and Aac(6') genes was detected, with 99.9% and 100% identity, in *P. aeruginosa* in 2004 (Genbank accession no.GQ302617, unpublished data) and *A. baumannii* in 2005, respectively (Kansakar *et al.*, 2011) (Appendix Figure A3.5). These isolates were from the same teaching hospital, Ramathibodi hospital, as MDR strains of *A. xylosoxidans*. Since strains R4 and R8 were collected in 2012 and 2010, respectively, this study suggested that this integron had been well conserved over eight years. In addition, the integron was normally merged into, and carried by, a plasmid, but this integron was inserted into the chromosome instead (Figure 5.11).

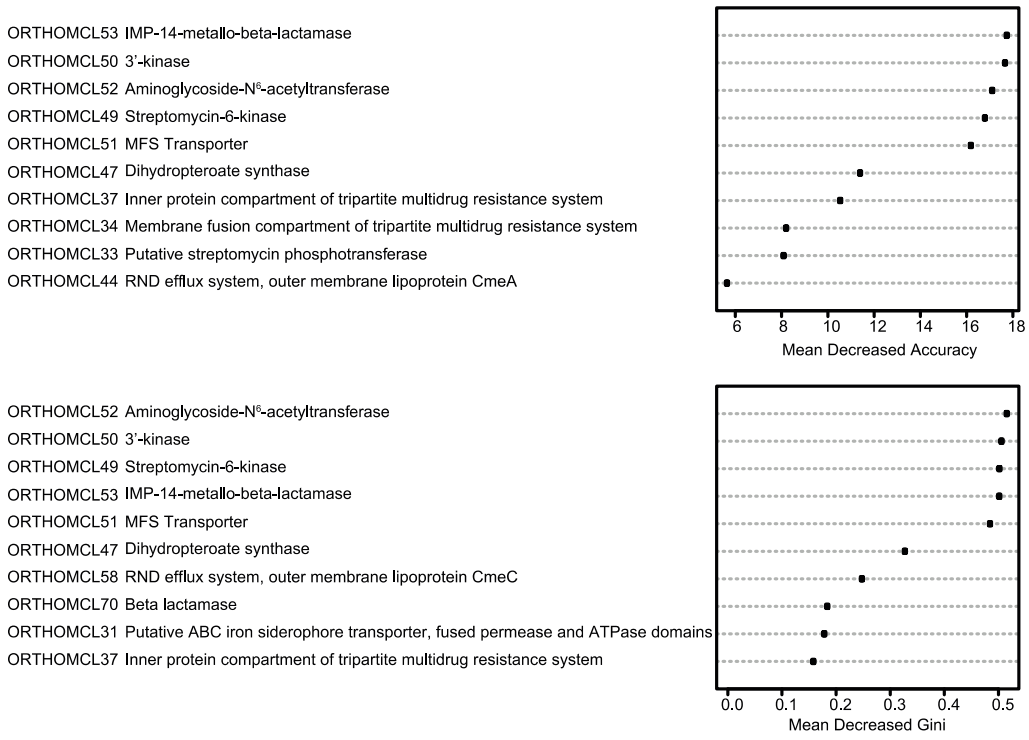


Figure 5.12: A variable importance plot based on Random Forest™ classification of carbapenem susceptibility in 25 isolates of *A. xylosoxidans*, showing the importance of each orthologue. The X-axes indicate the level of importance. The Y-axes show the list of orthologues.

Random Forest™ selected and ranked five orthologues: IMP-14-metallo-beta-lactamase, Streptomycin-6-kinase, Aminoglycoside-N^6-acetyltransferase, MFS permease (or Tetracycline resistance protein) and 3'-kinase (or Streptomycin resistance protein B) (Figure 5.12). These genes were present in two MDR strains from Thailand only. These orthologues had close mean decreased accuracy scores and mean decreased Gini scores. Interestingly, these five orthologues were also obtained from the analyses using susceptible profiles to other carbapenems. With the whole genome as an input, these orthologues were made up the top five rankings based on mean decreased importance scores and mean decreased Gini scores. Regarding the biological function of these genes, only IMP-14-metallo-beta-lactamase, which exhibits a carbapenemase function, contributed to carbapenem resistance, whereas the other genes had other functions related to antibiotic resistance. The physical relationships between IMP-14-metallo-beta-lactamase, Aminoglycoside-N^6-acetyltransferase and 3'-kinase (or Streptomycin resistance protein B) in those two Thai MDR strains (Figure 5.11) allowed us to ascertain the statistical relationships between the genes (Figure 5.12).

5.3.6. Other resistance-associated proteins

Apart from major resistance genes, efflux pumps and beta-lactamases, other enzymes related to drug resistance were identified using the protein BLAST against CARD and the NCBI database. As presented in Table 5.4, seven orthologous groups containing resistance-associated proteins were conserved across 26 genomes. Four out of these seven orthologous groups were enzymes that are normally present in prokaryotic organisms: dihydropteroate synthase, an enzyme in folic acid biosynthesis pathway, and topoisomerases II (DNA Gyrase) and IV - enzymes that relieve strain on the DNA string. The rest of the intrinsic resistance-associated enzymes were a fosmidomycin resistance protein, chloramphenicol acetyltransferase and aminoglycoside acetyltransferase. These could be directly involved in drug resistance due to their functions.

5.3.6.1. Dihydropteroate synthase

Dihydropteroate synthase is an enzyme occurring in the folic acid biosynthesis pathway, which is important in living organisms. It is a target of a sulfonamide drug called sulfamethoxazole (McCullough & Maren, 1973). The mutation of this enzyme causes bacteria to become resistant to sulfonamide antibiotics. Even though this gene was found to be conserved across *A. xylosoxidans* (Table 5.4), resistance to sulfamethoxazole was not detected in any of the *A. xylosoxidans* isolates (Table 5.2 and 5.3) except for DSM 2402, R4 and R8, which were less susceptible to sulfamethoxazole. Another dihydropteroate synthase, in addition to those found in the groups of intrinsic genes, was detected in the sulfamethoxazole-resistant strains.

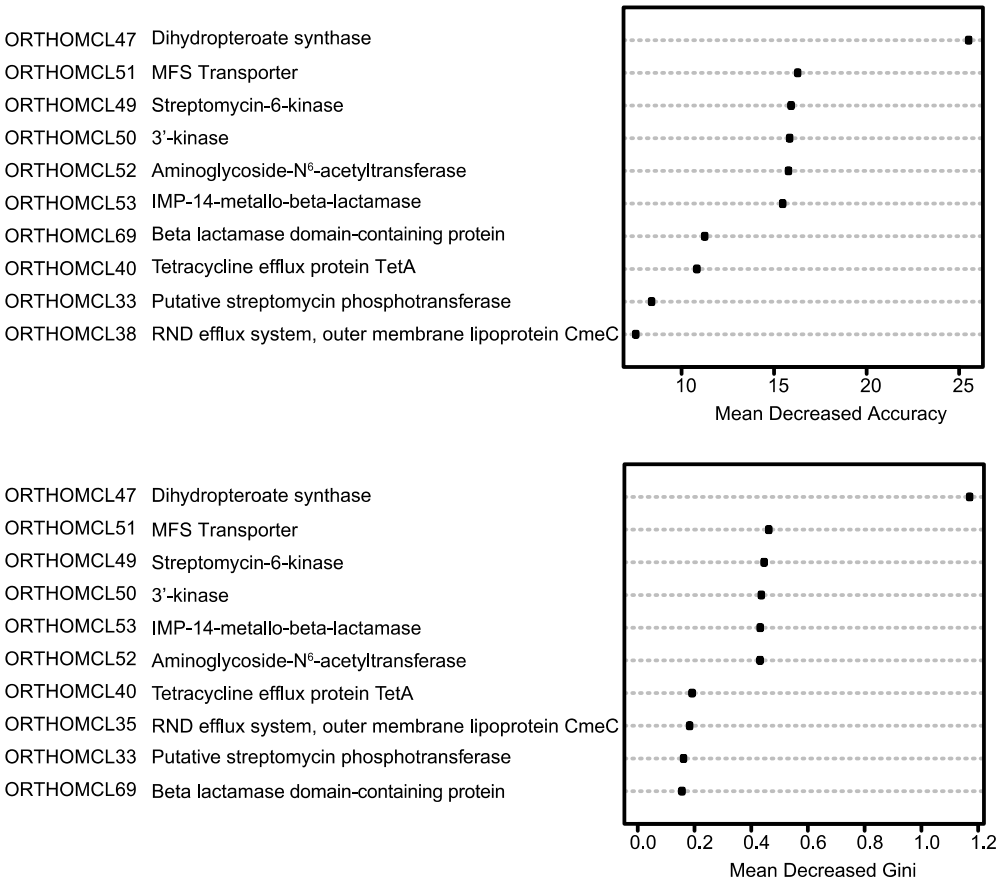


Figure 5.13: A variable importance plot based on Random Forest™ classification of sulfamethoxazole/trimethoprim susceptibility in 25 isolates of *A. xylosoxidans*, showing the importance of each orthologue. The X-axes indicate the level of importance. The Y-axes show the list of orthologues.

Random Forest™ ranked dihydropteroate synthase orthologues the first amongst 74 resistance gene candidates. As shown in Figure 5.13, this orthologue was found to have the highest mean decreased accuracy score (Score = 25.5) and mean decreased Gini score (Score = 1.17), which means that it is important in classification with high purity.

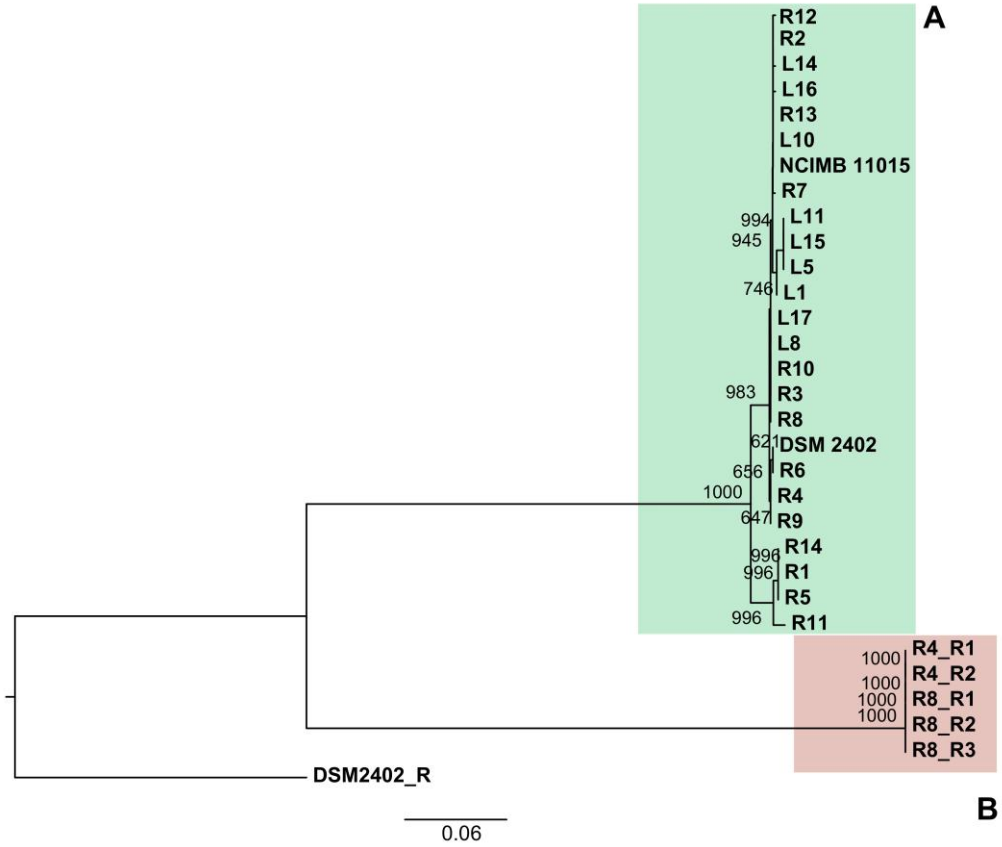


Figure 5.14: A Neighbor-Joining phylogenetic tree showing genetic relationships between dihydropteroate synthase genes in *A. xylosoxidans*. A green box indicates housekeeping dihydropteroate synthase genes, which are important in folic acid biosynthesis in bacteria. An orange box indicates resistance-related dihydropteroate synthase genes, which are identified in SXT-resistant strains. NH44784-1996 stands for *A. xylosoxidans* NH44784-1996, NCIMB 11015 for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; L-strains are Liverpool's clinical isolates; and R-strains are Thai clinical isolates. The suffix 'R' signifies a gene that contributes towards trimethoprim resistance. Bootstrap values more than 50% were indicated.

As shown in Figure 5.14, a Neighbor-Joining phylogenetic tree of dihydropteroate synthase genes found in 26 *A. xylosoxidans* genomes illustrated two main clades of genes: clade A (highlighted in green) and clade B (highlighted in orange). Clade A consisted of dihydropteroate synthase genes from core genomes, whereas clade B consisted of genes from non-core genomes, which represents lateral transfer of the genes. The presence of resistance-associated dihydropteroate synthases (clade B in Figure 5.14) was related to SXT resistance in those strains. It could be implied that the resistance to sulfamethoxazole in *A. xylosoxidans* in this study was not caused by the mutation of conserved dihydropteroate synthase, but by resistance genes acquired from external sources.

5.3.6.2. Topoisomerase

Topoisomerase is a group of enzymes that unwind DNA strands so as to reduce strain on the strands when the DNA strings are taken apart in replication process. Since the enzyme is a target of fluoroquinolone antibiotics (Aldred, Kerns & Osheroff, 2014), a mutation in the enzyme is considered one of the fluoroquinolone resistance mechanisms. The fluoroquinolone resistance-determining region (QRDR) is a region of the enzyme that interacts with fluoroquinolone. The QRDR is determined by two amino acid residues, Serine at position 83 and Aspartic acid at position 87 in topoisomerase II subunit A. Several studies have shown that mutations mostly happen near the QRDR (Heddle & Maxwell, 2002; Aldred, Kerns & Osheroff, 2014), as shown in Table 5.6. This region is a part of a binding pocket for both DNA and quinolone antibiotics.

Table 5.6: Amino acid substitutions in subunit A of the topoisomerase II enzyme, which is associated with fluoroquinolone resistance (Friedman, Lu & Drlica, 2001)**.**

Amino acid position	Native	Substitute
67	Alanine (A)	Serine (S)
51	Alanine (A)	Valine (V)
81	Glycine (G)	Cysteine (C)
82	Cysteine (C)	Alanine (A)
83*	Serine (S)	Leucine (L)
83*	Serine (S)	Tryptophan (W)
84	Alanine (A)	Proline (P)
87*	Aspartate (D)	Asparagine (N)
87*	Aspartate (D)	Tyrosine (Y)
106	Glutamine (Q)	Histidine (H)

* Quinolone resistance-determining region position

As seen in Table 5.2 and Table 5.3, the strains were less susceptible to fluoroquinolone due to the presence of either intrinsic mutations or other genes that cause resistance. As described in the section entitled, 'Efflux pump', the presence of intrinsic RND-type efflux pumps that are associated with fluoroquinolone resistance in the genome of *A. xylosoxidans* could explain an innate resistance in this species (Bador *et al.*, 2013).

Here, all genes encoding topoisomerase IV, except for topoisomerase IV subunit A, was found as conserved resistance genes. The loss of topoisomerase IV subunit A in strains R9 and R10 was probably due to an error in sequencing or a natural absence but it did affect their susceptibility to fluoroquinolone. A multiple sequence alignment was performed using genes encoding subunit A of topoisomerase II in *A. xylosoxidans* (Appendix Figure A3.6A), a region which commonly determines fluoroquinolone resistance in Gram-negative bacteria (Redgrave, 2014). It showed no significant amino acid substitutions that lead to fluoroquinolone resistance (Hallett & Maxwell, 1991; Jalal *et al.*, 2000), except in R8 (marked by an asterisk in Appendix Figure A3.6A). The substitution involved the replacement of Aspartic acid (D) with Asparagine (N) at amino acid position 87, which is a part of the QRDR. As expected, the mutation of the

QRDR changed the polarity of a binding site (from acidic to basic), which could affect the binding of fluoroquinolone in the pocket. It finally resulted in a remarkably high MIC of LEV in R8 (Appendix Table A3.4). Even though the fluoroquinolons resistance was described by the presence of the efflux pump, AxyABM, this demonstrated that the mutation in topoisomerase II contributed towards a more resistant phenotype with higher amount of fluoroquinolone.

5.3.6.3. Modifying enzymes

Chloramphenicol acetyltransferase is an enzyme that inactivates the chloramphenicol antibiotic by adding an acetyl group to the chloramphenicol molecule (Murray, 1997). This enzyme was detected as one of the core resistance genes in this analysis (Table 5.4). The identification of chloramphenicol acetyltransferase genes in *A. xylosoxidans* was reported in the complete genome of *A. xylosoxidans* strain NH44784-1996, which had a chloramphenicol-resistant phenotype (Jakobsen *et al.*, 2013). On the other hand, the gene was also detected in chloramphenicol-susceptible strains (Hu *et al.*, 2015).

Aminoglycoside acetyltransferase is an enzyme that modifies the structure of aminoglycosides by adding acetyl groups to aminoglycosides. This modification results in an activation of the antibiotics, which makes the bacteria resistant to aminoglycosides (Ramirez & Tolmasky, 2010). Orthologues of the aminoglycoside-N^3-acetyltransferase were conserved across 26 genomes. Therefore, the finding of this conserved gene, together with the expression of efflux pumps previously mentioned on the genomes suggests an innate resistance to aminoglycosides in *A. xylosoxidans*. Moreover, a novel aminoglycoside acetyltransferase, aminoglycoside-N^6-acetyltransferase, was detected in multidrug-resistant (MDR) *A. xylosoxidans* strains, R4 and R8. After further investigation into the genomes of the MDR strains, the aminoglycoside-N^6-acetyltransferase was located on their genomes. It was found to be surrounded by genes such as integrase and conserved sequences (a diagram of the gene cassette is shown in Figure 5.11) such as *attI* and *attC* sites, which are generally found in the integron. Several studies have revealed that a gene encoding aminoglycoside

acetyltransferase was a part of a resistant gene-carrying integron (Zhao & Hu, 2011; Traglia *et al.*, 2012; Di Pilato, Pollini & Rossolini, 2014; Touati *et al.*, 2013; Chen *et al.*, 2014).

5.4. Discussion

Antibiotic resistance has been considered a major problem due to the difficulty to select antibiotics to eliminate those resistant bacteria. *A. xylosoxidans* is a bacterium that is famous for its multi-drug resistance (Hu *et al.*, 2015). A study on antibiotic resistance in *A. xylosoxidans*, looking at both phenotypic attributes and genetic attributes, revealed both intrinsic resistance and acquired resistance on its chromosome. In addition, a comparative study of the resistance phenotype and the genome of *A. xylosoxidans* identified acquired resistance that is regulated by mobile genetic elements.

Table 5.7: A summary table presenting drug resistance mechanisms identified in *A. xylosoxidans* in this study (categorised by resistance types - intrinsic and acquired)

	Groups of antibiotic	Possible mechanism of resistance
Intrinsic	Aminopenicillin Cephalosporin	- CmeABC (AxyABM homologue) - Beta lactamase
	Aminoglycoside	- CmeABC (AxyXY-OprZ homologue) - Aminoglycoside-N^3-acetyltransferase
	Fluoroquinolone	- CmeABC (both AxyABM and AxyXY-OprZ homologue) - Mutation of QRDR of the subunit A of DNA gyrase
Acquired	Aminoglycoside	- Aminoglycoside-N^6-acetyltransferase
	Carbapenem	- IMP-14 metallo-beta-lactamase
	Aminopenicillin	- Class A beta-lactamase
	Folate inhibitor	- Drug resistant dihydropteroate synthase

5.4.1. Intrinsic antibiotic resistance and associated genes

Innate resistance in *A. xylosoxidans* was characterised by the presence of an OXA-114 beta-lactamase (Doi *et al.*, 2008) and an RND-type efflux pump (Bador *et al.*, 2011, 2013). OXA-114 beta-lactamase is the only beta-lactamase that has been identified and well characterised. In this study, seven orthologues

of a commonly shared beta-lactamase were identified across 26 *A. xylosoxidans* genomes (Table 5.4). This agrees, to a certain degree, with previous studies, which reported that there were at least five beta-lactamase enzymes found in *A. xylosoxidans* ATCC 27061 (Hu *et al.*, 2015) and ten beta-lactamases found in *A. xylosoxidans* NH44784-1996 (Jakobsen *et al.*, 2013). It is believed that these beta-lactamases mainly contributed to the resistance to beta-lactam in these *A. xylosoxidans* strains (Table 5.7).

As well as beta-lactamase genes, a number of efflux pumps were identified as intrinsic resistance genes. There are five families of efflux pumps associated with antibiotic resistance mechanisms in bacteria (Webber, 2002): major facilitator, multidrug and toxic effect, RND, small multidrug resistance and ATP-binding cassette. Comparative genomics of resistance genes in genus *Achromobacter* revealed that the number of MDR pumps ranges from 40 to 53 genes, depending on the species. Despite the fact that MDR efflux systems in genus *Achromobacter* have been studied, the RND-type pump is the only efflux pump that is well identified, including AxyABM and AxyXY-OprZ (Bador *et al.*, 2011, 2013). These pumps are associated with resistance to fluoroquinolones, cephalosporins and aminoglycosides (Table 5.7). In the same manner, 45 resistance pump-related genes were discovered in the complete genome of *A. xylosoxidans* NH44784-1996 (Jakobsen *et al.*, 2013) and 38 resistance pump-related genes were identified in the genome of *A. xylosoxidans* ATCC 27061 (Hu *et al.*, 2015). In keeping with previous studies, RND-type efflux pumps were the largest groups of genes found in *A. xylosoxidans* in this study (see Chapter 4). With orthologous grouping and resistance gene re-annotation, 18 common drug resistance-related families were detected across 26 *A. xylosoxidans* genomes (Table 5.4).

Our evidence that AxyXY-OprZ homologues are absent in aminoglycoside-susceptible *Achromobacter* species (Figure 5.9) strongly supports the theory that an AxyXY-OprZ homologue contributes towards aminoglycoside resistance in *A. xylosoxidans*. Consequently, it can be assumed that the presence of these intrinsic efflux transport systems may affect the innate resistance of multiple antibiotics (Table 5.7). An RND-type efflux transport system contains three main subunits:

an inner membrane protein (for example, CmeB and AxyY), an outer membrane-integrated channel (such as CmeC and OprZ), and an adaptor/accessory protein (for example, CmeA and AxyX) (Nikaido, 2011). The inner membrane protein is the most essential part of the system because this subunit plays a role in binding drug molecules and expelling them from the cell. The adaptor protein has a minimal role in drug resistance. Nevertheless, evidence shows that different adaptor proteins affect the compatibility of the drug and the inner membrane protein (Krishnamoorthy, Tikhonova & Zgurskaya, 2008). The last part of the system is the outer membrane channel. The channel provides the gateway for the expelled molecule (Blair & Piddock, 2009). Comparative analysis of the structure of the channel shows that the different channel molecules have similar gross structures whilst their sequences are different (Federici *et al.*, 2005). As mentioned above, the inner membrane protein is, therefore, crucial to the resistance mechanism of the RND pump system.

The innate presence of drug resistance enzyme modification enhances drug resistance ability in the strains. In addition to expressing CmeABC efflux transport systems that pump out aminoglycosides, *A. xylosoxidans* have an aminoglycoside-N^3-acetyltransferase (Table 5.7), which confers aminoglycoside resistance (Davies & O'Connor, 1978). However, this conserved acetyltransferase enzyme probably has a non-significant effect on aminoglycoside resistance in *A. xylosoxidans*, compared to the efflux pump. This is demonstrated by the existence of strains that lack an AxyXY-OprZ homologue pump and are aminoglycoside-susceptible despite having an acetyltransferase in their genome.

5.4.2. Acquired resistance

In addition to core resistance genes, acquired resistance-associated genes were also identified in the genomes of 2 Thai MDR *A. xylosoxidans* strains. As shown in Table 5.2 and Table 5.3, resistance to carbapenems was unique to strain R4 and strain R8. Yamamoto *et al.* (2012) have reported carbapenem-resistance in *A. xylosoxidans* possessing a class 1 integron that carries IMP-1 and IMP-19

metallo-beta-lactamase. This study is the first to report the detection of a class 1 integron carrying IMP-14 metallo-beta-lactamase from Thai strains of *A. xylosoxidans* (Table 5.7). IMP-14 was detected in Thai clinical strains of *P. aeruginosa* (Genbank accession no. GQ302617, unpublished data) and *A. baumannii* in 2005 (Kansakar *et al.*, 2011). These two species were clinical isolates from the same hospital from whence two MDR *A. xylosoxidans* isolates were collected. This suggests the appearance of horizontal gene transfer in this species.

When compared to the gene cassette found in *P. aeruginosa* and *A. baumannii* (Appendix Figure A3.5), the structure of the integron - including an integrase, an aminoglycoside-N^6-acetyltransferase, a sulfamethoxazole-resistance dihydropteroate synthetase (Table 5.7) and a conserved sequence - demonstrates high percentage similarity (99.9% and 100%). This evidences the local transmission and the preservation of the genetic element across the species. Interestingly, the integron was intact on the chromosome, which is consistent with the finding that *A. xylosoxidans* can harbour mobile genetic elements on its chromosome (Hu *et al.*, 2015). Taken together with the fact that, as discussed in Chapter 4, the *A. xylosoxidans* genome is open, this may shed the light on why the genome of *A. xylosoxidans* is likely to be receivable and likely to be a reservoir of resistance genes.

5.4.3. Beta-lactam resistance

In general, *A. xylosoxidans* is recognised as a species that is resistant to several antibiotics, especially beta-lactam antibiotics. The mechanism for developing beta-lactam resistance could be the production of beta-lactamase, the expression of efflux transporter and the production of modified drug targets. The presence of a number of beta-lactamase genes was shown by previous studies (Jakobsen *et al.*, 2013; Hu *et al.*, 2015). The discovery of an intrinsic class D beta-lactamase, oxacillinase-114, in the species proved that beta-lactam resistance in *A. xylosoxidans* is innate and gene expression is not induced (Doi *et al.*, 2008). Moreover, the presence of a metallo-beta-lactamase is acquired and transferred

via a genetic element (Sofianou *et al.*, 2005; Neuwirth *et al.*, 2006; Di Pilato, Pollini & Rossolini, 2014). In the comparative analysis in this study, the IMP-14-metallo-beta-lactamase was revealed in the strains with less susceptible to carbapenems (Table 5.2, Table 5.3, Figure 5.5 and Figure 5.6). The presence of a class A beta-lactamase could explain the resistance to aminopenicillin in *A. xylosoxidans* (Table 5.7).

In addition, an efflux pump associated with resistance to cephalosporin was indicated in the genome of *A. xylosoxidans*. Inactivation of these two RND-type efflux pump genes, AxyABM and AxyXY-OprZ, resulted in a decrease in cephalosporin MIC values (Bador *et al.*, 2011, 2013). In agreement with previous studies, RND-type efflux pump genes were also identified as intrinsic resistance genes of the strains (Table 5.4). These findings suggested that the beta-lactam resistance mechanism in *A. xylosoxidans* would be due to both beta-lactamases and efflux pumps.

5.4.4. Aminoglycoside resistance

Aminoglycoside is an antibiotic that is commonly used against various pathogens, including Gram-negative bacteria. *A. xylosoxidans* has been shown to be aminoglycoside resistant (Yabuuchi & Oyama, 1971; Pien & Higa, 1978; Shigeta *et al.*, 1978; Giacoia, 1990; Bador *et al.*, 2011, 2013). The mechanisms leading to aminoglycoside resistance are the modification of the drug by enzymes, the transportation of the drug out of a cell by a pump, and the modification of the drug target. An RND-type efflux transporter, AxyXY-OprZ (Table 5.7), is the only aminoglycoside resistance-associated protein that has been well characterised and identified in *Achromobacter* species (Bador *et al.*, 2013). The analysis in this study also presented that an AxyXY-OprZ homologous pump could not be identified in aminoglycoside-susceptible species. Moreover, the presence of AxyXY-OprZ homologous CmeABC in less-aminoglycoside susceptible strains supported the evidence that the efflux pump was strongly associated with an aminoglycoside-resistant phenotype. This leads to the implication of using these genes as a screening test for aminoglycoside resistance.

A drug-modifying enzyme, aminoglycoside acetyltransferase, is an enzyme believed to have a role in aminoglycoside resistance by structurally inactivating agents (Wright, 1999). It has been reported that the existence of this modifying enzyme is associated with aminoglycoside resistance in *E. coli* and *Klebsiella spp.* (Haldorsen *et al.*, 2014). Isolates used in this study contain this enzyme in their core genome and MDR isolates contain an extra gene encoding aminoglycoside acetlyltransferase on an integron (Table 5.7, Figure 5.11). The comprehensive analysis showed that a CmeABC pump, with a high identity to AxyXY-OprZ, and an aminoglycoside acetyltransferase were present in the intrinsic resistance gene set of isolates, suggesting that aminoglycoside resistance is an innate ability of *A. xylosoxidans*.

5.4.5. Fluoroquinolone resistance

Agents in the fluoroquinolone class are some of the most widely clinically-used antibiotics due to their broad spectrum. Bador *et al.* (Bador *et al.*, 2011) has described AxyABM as an efflux pump that has a role in multidrug resistance in *A. xylosoxidans* AXX-A. This is the only mechanism that has been identified that can explain the fluoroquinolone resistance mechanism in genus *Achromobacter*. Genes that have close homology to the efflux system's genes were identified as intrinsic resistance genes of *A. xylosoxidans*.

The mutation of DNA Topoisomerase effects a massive change in MIC (Redgrave *et al.*, 2014) in *E. coli*. In Gram-negative bacteria, the mutation in the QRDR in DNA Topoisomerase subunit A is used to indicate resistance to fluroquinolone (Heddle & Maxwell, 2002). In this study, the mutation in the QRDR region (Appendix Figure A3.6) explains the higher resistance to fluoroquinolones seen in *A. xylosoxidans* strain R8.

5.4.6. Sulfamethoxazole-trimethoprim resistance

Sulfamethoxazole-trimethoprim is a combination of agents that is potent in treating a broad range of pathogens, especially in urinary tract infections or in systemic infections (Huovinen *et al.*, 1995; Eliopoulos & Huovinen, 2001). Due to the mechanism of action and spectrum of activity, this combination is extensively used as a first line drug against urinary tract infections. For *Achromobacter* infections, this agent is usually effective against the pathogen (Derber *et al.*, 2011; Atalay *et al.*, 2012; Sheng *et al.*, 2013; Tena *et al.*, 2014). Resistance to SXT occurs mostly due to mutations in dihydropteroate synthase and dihydrofolate reductase, respectively. In agreement with previous studies, two groups of dihydropteroate synthase were identified here (Figure 5.14). Physically, this gene locates on the class 1 integron that is carrying IMP-14-metallo-beta-lactamase (Figure 5.11 and Figure 5.12). Some studies have shown sulfamethoxazole resistance to be passed between strains and species in a horizontal transfer event via a class 1 integron (Antunes *et al.*, 2005; Blahna, 2006). Altogether, it is reasonable to propose that sulfamethoxazole-trimethoprim resistance is acquired in *A. xylosoxidans* because of its open pan-genome (as discussed in Chapter 4).

5.5. Conclusions and future work

Here, a present study has illustrated the use of whole genome sequencing and bioinformatics tools to predict and reveal antibiotic resistance genes associated with a susceptibility profile. The susceptibility of isolates of *A. xylosoxidans* reveals that resistance to aminoglycosides, fluoroquinolones and cephalosporins are inherent phenotypes in *A. xylosoxidans*, as indicated by the presence of RND-type efflux pump genes and beta-lactamase genes. Fluoroquinolone resistance could be explained by the presence of a fluoroquinolone-associated efflux pump. Abnormally high resistance (high MIC) to fluoroquinolones is caused by a mutation of the QRDR. On account of the presence of the AxyXY-OprZ homologous efflux system in strains less susceptible to aminoglycosides, the presence of this pump, especially its inner membrane protein, could potentially be used as a screening marker for aminoglycoside resistance.

The pump system genes and beta-lactamase genes were included in accessory genes. Additionally, an extra metallo-beta-lactamase gene and resistant dihydropteroate synthase gene were found on a class 1 integron that has firmly merged into a chromosome of an open-genome species. According to the structure of the class 1 integron, the co-resistance to trimethoprim, aminoglycosides, and beta-lactamases could have lead to the presence of the integron in the isolates.

Using extensive bioinformatics analysis, Random Forest™ was able to select carbapenemase genes. The detection of these genes led us to a further investigation and the discovery of the chromosomally-lodged integron. This demonstrated the efficiency of an unsupervised-learning machine in linking biological information to computationally predicted information.

Chapter 6

Identifying virulence genes used by *Achromobacter xylosoxidans* in an infection model

6.1. Introduction

6.1.1. Virulence in bacterial infection

In the environment bacteria are often found as part of bacterial communities and the human body is no exception. There are many examples of bacterial communities associated with humans, which are referred to a microbiomes. The relationship between microbiomes and their hosts can be often defined as 'commensalism', where the microbes gain benefit from the host without affecting the host's fitness. Interactions can also be 'mutualism', where both host and microbes gain benefit from each other, as exemplified by gut microbes that synthesise vitamin for their host (LeBlanc *et al.*, 2013). Conversely, there are also pathogenic relationships where microbes detrimentally affect the fitness of the host.

Bacteria are evolving in the contact with the host and both partners can be expected to evolve strategies for co-existence. This can result in an "arms race" between the host and the microbe as each tries to overcome the others defences or infection strategies (Brown *et al.*, 2006). If the bacteria can overcome host defence, it is considered virulent colonisation of the bacteria. By etymology, virulence derives from Latin words *virulentus* (full of poison): *virus* (poison) and *lentus* (fullness) (Casadevall & Pirofski, 2001). Virulence is also defined as 'the capacity of a microorganism to infect and damage a host' (Lipsitch & Moxon, 1997). The important mechanisms of virulence are toxicity, attachment, aggressiveness and immunity avoidance (Casadevall & Pirofski, 2001).

Toxicity is determined as the ability to produce a substance that causes damage to host tissue. The substance is called a 'toxin'. The toxin can be categorised into two main groups: endotoxin and exotoxin. Endotoxin is a structure that is commonly present as a part of bacterial cell wall and is released after the death of the microbes. Lipopolysaccharide (LPS) is the endotoxin of Gram-negative bacteria and techoic acid and peptidoglycan are the endotoxin of Gram-positive bacteria. The molecules are recognised by toll-like receptor (TLR) moleules, which are present in most of innate immune cells. LPS is recognised by TLR-4 (Erridge, Bennett-Guerrero & Poxton, 2002), whereas Techoic acid and peptidoglycan are recognised by TLR-2 (Schwandner *et al.*, 1999). This leads to the production of inflammatory cytokines (e.g. interleukin-1 and interleukin-12) that orchestrate inflammatory response in host (Miller, Ernst & Bader, 2005). Inversely, exotoxin, sometimes called 'true toxin', is defined as the toxic substance that is secreted by a living microorganism (Casadevall & Pirofski, 2001). The effect of exotoxin leads to several detrimental infections; for example, Cholera toxin of *V. cholerae* and Staphylococcal toxic shock syndrome toxin-1 of *S. aureus*.

Attachment is also considered as a characteristic of virulence. The attachment of bacteria to host tissues, in particular, epithelial surface, is an essential process for some species to cause disease. For example, the firm adherence of uropathogenic *E. coli* to urinary tract using P and type I fimbria protects the bacteria from shedding by urine flow (Melican *et al.*, 2011). This gives an advantage to bacteria to remain their colonisation, leading to the inflammation of urinary tract system. Nevertheless, not all bacteria attached to host's tissue are virulence. Commensal bacteria in human gastrointestinal tract adhere to mucosal surface as a protective layer to prevent the attachment of non-commensal bacteria (Tlaskalová-Hogenová *et al.*, 2011).

Aggressiveness is the ability to invade and reproduce in the host. This attribute of virulence refers to any mechanism that promotes bacterial survival in the host and, subsequently, leads to the damage to host (Casadevall & Pirofski, 2001). Most of the mechanisms that contribute to the aggressiveness of the bacteria are

to invade or avoid host's immune response. The production of polysaccharide capsule is the example of immune avoidance mechanism. The polysaccharide capsule produced by some species, such as *S. pneumoniae*, inhibits host's immunity by means of inhibiting complements and phagocytosis (Hyams *et al.*, 2010). This helps bacteria to survive and, subsequently, damage the host using other virulence factors, such as hyaluronidase in *S. pneumoniae* (Kadioglu *et al.*, 2008).

One important parameter determining the pathogenicity of the bacteria is the ability to grow and to reproduce in the host. It is called a 'fitness' of pathogen. Both host and pathogen try to retain their fitness by overcoming one another. For example, host increases body's temperature to suppress the reproduction of the bacteria and the bacteria try to avoid thermoregulator by rapidly killing the host (Elliot, Blanford & Thomas, 2002). Therefore, the pathogen needs to produce the factor that makes them dominate host's fitness.

Virulence factor is defined as the factor that helps bacteria to retain their virulence, and the virulence is weakened when the factor is diminished (Casadevall & Pirofski, 2000). According to the attributes of bacterial virulence described above, the virulence factor, therefore, promotes those abilities in the host. Nonetheless, the concept of virulence of the pathogen is dependent on the susceptibility of the host. Therefore, this defines a term 'pathogen' as an organism that can colonise a host and can cause diseases (Brown *et al.*, 2006).

The genes used by *A. xylosoxidans* in pathogenic infection are currently unknown. There is limited knowledge of the virulence factor produced by *A. xylosoxidans*. It has been recognised that the bacteria is opportunistic infection, determined by *in vivo* experiment (Hyodo, Katahira & Shigeta, 1982), and the bacteria secretes heat-stable inflammatory factor, determined by *in vitro* experiment (Mantovani *et al.*, 2010).

6.1.2. Infection models

Animal models of infection are important tools for experimental immunologists and microbiologists to understand defense mechanisms of immune systems and how pathogens avoid the host's immune system. Whole organism models include several levels of organisation and complexity e.g. organisms, organs, tissues and culture. According to different level of infection model, the whole body can provide a better understanding of how a pathogen behave during infections and cause disease. Moreover, organisms with close relationship to human can demonstrate host response better. For human diseases, mammal models such as mouse, pig, rabbit and ape, are relevant as their genetics, physiology and immune systems can have direct parallels humans (Hedges, 2002; Hunter, 2008). In addition, genetic manipulation and engineering of mammals can provide researchers with the genetic tools to understand the role of genes with a human homolog.

The mouse model is a very important for scientific research, including the study of infectious diseases. Mice share about 70% of amino acids sequences with humans (Emes, 2003), so molecular mechanisms in mice often have human homologs and *vice versa*. A number of studies to understand human immune responses have been conducted in mice (Buer & Balling, 2003). Despite similarities between mice and humans it is often necessary to use genetic modification to determine particular process. There are a number of mutant mice available to support study in immunity and infection.

Prior to the utilisation of animal infection model, the principle of 3Rs should be considered. The 3Rs are the framework developed for animal use in scientific research. 3Rs consist of (1) replacement, (2) reduction, and (3) refinement (https://www.nc3rs.org.uk/the-3rs). Firstly, replacement means to avoid animal use or to replace vertebrate animal model with another model, including human tissue, established cell line, immature animal, and computer model. An example of replacement in infection model is the replacement of mouse by *G. mellonella* in *P. aeruginosa* infection (Jander, Rahme & Ausubel, 2000). Secondly, reduction refers to reduce the number of animals used in the experiment. This can be done by comprehensively designed experiments, sharing information between groups, and using advanced data collection technology. For example,

green fluorescent protein-tagged bacteria can be used to show both mortality of the animal and *in vivo* distribution of the bacteria in a single experiment (Park *et al.*, 2012). Lastly, refinement is important in the studies that the use of animal model is unavoidable. This issue involves animal husbandry and welfare, which affect the quality of life of animal (Morton, 1998). Good welfare allows animals not to be stressed, resulting in interpretable result with a reduced number of animals used (Baker *et al.*, 2011). Above all, cost effectiveness of the model is an issue to be considered. Mammalian models are more expensive than non-mammalian model, however, mammalian models can provide researchers with more comprehensive data. The 3Rs framework provides the researchers with guideline and ethical issue to consider the use of animal model in the scientific experiment.

Following the principle of 3Rs, if the use of animal model is unavaoidable, the replacement is regarded as the most important issue. To investigate the pathogenicity of the pathogen, systemic infection model, such as animal, remains important. The replacement of mammalian model by non-mammalian model is under consideration. Non-mammalian models are an important alternative for infection and immunology. These models can be used directly as infection models, but also for initial screens of large bacterial strain panels, prior to performing smaller scale mammalian model tests. Non-mammalians are non-sentient organisms. Furthermore, they have demonstrated the potential to replace mammalian model in various studies. Non-mammals models include, the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster* and the zebrafish *Danio rerio*.

C. elegans was the first non-mammalian model used to study infectious diseases. It has often been used as a test for virulence in clinical strains and in screening tests for antibiotic treatments. Owing to lack of adaptive immune response, the nematode provides a good model for observing innate immune responses. For example, in *E. faecalis* infection, the immune response of the nematodes demonstrated the importance of reactive oxygen species as a defence mechanism (Chávez *et al.*, 2007), which are similar to the response shown by human phagocytic cells.

The fruit fly is a well-known model for human genetics and cancer studies. Based on its behavior, the fly consumes rotten fruit, which contains many bacteria, so the fly produces number of antimicrobial agent into the gut in order to defend against those bacteria. In addition, the anatomy of the gut and immune signaling pathways have some conserved features with humans, so *D. melanogaster* can be used as a model for human gut infection (Apidianakis & Rahme, 2011). Flies have also been used to study Gram-positive bacterial infections, such as *S. aureus* (Needham *et al.*, 2004) due to the presence of Peptidoglycan-specific Pattern-recognition Receptor (Pili-Floury *et al.*, 2004).

The zebrafish is an important model for the study of embryology in medical research. The zebrafish has both innate and adaptive immune systems. For innate immunity, the phagocytes and, interestingly, Toll-like receptors play a role similar to immune response in human (Phelan, Mellon & Kim, 2005). Enzymes involved in adaptive immune system (Willett, Cherry & Steiner, 1997), as well as B cells (Danilova & Steiner, 2002) and T cells (Danilova *et al.*, 2004), were found in the fish. Humoral immune response is also conserved between zebrafish and human. With exposure to pathogens, inflammatory cytokines, such as interleukin-1 and tumor-necrosis factor are produced to respond to infection in zebrafish (Pressley *et al.*, 2005). It is agreed that the zebrafish is an effective model for infection and immunology study.

All of human pathogens are able to grow well at 37 °C, which is the physiological temperature in human. However, optimal temperature of the all aforementioned organisms are below 37 °C. Substituted pathogens are often used to study the course of infection in those models, as exemplified by the use of *M. marinarum* in the Zebra fish to imitate human tuberculosis from *M. tuberculosis* (Meeker & Trede, 2008). Non-mammalian model being able to grow at 37 °C is considered an appropriate model to demonstrate *in vivo* infection.

6.1.3. The use of *G. mellonella* as an infection model in infection biology

The wax moth larvae, *G. mellonella*, provides experimental biologists with a convenient model, compared to other models. There are a number of advantages in using *G. mellonella* as an infection model.

1. The larvae is relatively large (2-4 cm). The size of animal is large enough for a researcher to hold and perform the experiment: For example, bacteria can be injected into the larvae via its left proleg to mimic bacteraemia, and antimicrobial injected to test drug's efficacy.
2. Haemolymph of the worm contains phagocytes, which its action is similar to human phagocytes, such as they produce radical oxygen species to combat foreign organisms (Bergin *et al.*, 2005).
3. The worm can be cultured at 37°C. This is important when considering its use as model for human infection, because human pathogens can be grown at 37 °C.

The use of this larvae as an infection model has been increasing in recent years. Many important human pathogens have been tested using *G. mellonella* model. A study in *P. aeruginosa* illustrated a power of *G. mellonella* that is positively correlated with *P. aeruginosa* experiment in mice (Jander, Rahme & Ausubel, 2000). Virulence of different strains were tested in *B. cepacia* complex (Seed & Dennis, 2008), *E. faecium* (Lebreton *et al.*, 2011) and Uropathogenic *E. coli* (Alghoribi *et al.*, 2014). Moreover, *G. mellonella* has also been used to reveal the efficacy of antibiotic against pathogens, such as *A. baumannii* (Peleg *et al.*, 2009; Hornsey *et al.*, 2013) and *B. pseudomallei* (Thomas *et al.*, 2013). These provide evidence that *G. mellonella* is an appropriate infection model and an alternative to a mouse model, when considering the 3Rs.

6.1.4. Aims and objectives

A. xylosoxidans is often described in a clinical context as an opportunistic pathogen, and a clinical strain has been shown to be lethal in immune-compromised mice (Hyodo, Katahira & Shigeta, 1982). However, there is no information about the relative virulence of *A. xylosoxidans* strains and no candidate virulence genes have been described.

The aim of this study is to investigate whether the larvae of the great moth (*G. mellonella*) could be used as an infection model for *A. xylosoxidans*. To specifically investigate whether there is difference between virulence determinants of the clinical strains (e.g. DSM 2402) and virulence determinants of environmental strains (e.g. NCIMB 11015). In addition to assess the relative virulence of *A. xylosoxidans* strains and use Random Forest™ data mining approach to determine the genetic bases of the any observed differences in virulence.

6.2. Material and methods

In order to investigate the virulence amongst *A. xylosoxidans* isolates, infections were performed in the *G. mellonella* model.

6.2.1. Determination of LD_{50} of type strains, NCIMB 11015 and DSM 2402, in *G. mellonella* infection model

Firstly, the difference in virulence between environmental strain, NCIMB 11015, and clinical lab strain, DSM 2402, were determined using lethal dose -50 (LD_{50}) of each strain. Prior to experiment, larvae were divided into four groups; two controls: non-injection, mock infection, and two experimental infections of strains NCIMB 11015 and DSM 2402. Each group contains 32 worms and was replicated three times. Overnight LB cultures of the NCIMB 11015 and DSM 2402 strains, were inoculated into LB broth, grown until they reached an optical density at 600 nm of 0.3 to 0.4 (approximately $1x10^8$ cells/ml). Then the cultures were centrifuged at maximum speed and resuspended in phosphate buffer saline (PBS) twice prior to inoculating into larvae. This was used as a stock inoculum, approximately $1x10^8$ cells/ml, to setup dilutions of cells for the experiments.

Randomly selected larvae were injected with 10 µl of previously prepared inocula using a 50 – µl Hamilton syringe. The injection was achieved via the last left proleg of the larvae. The larvae were injected with three different amounts of bacteria, 10^4 cells, 10^5 cells and 10^6 cells of each bacterial strain. After injection, the larvae were kept in 24-well plates (one larva per well) and were incubated at 37°C. Two control groups were included in this experiment, which were larvae injected with PBS (mock infection) and larvae with no injection in order to assure that larvae did not die of trauma from injection. The number of dead larvae was scored every 12 hours for 96 hours. Larvae were considered 'death' when they did not respond to physical stimulation and the physical appearance became melanised (Peleg *et al.*, 2009). Survival of larvae was assessed using Kaplan-Meier survival curves and Log-rank test in *survival* package in R version

3.1.2. Finally, The LD_{50} was calculated from a regression plot between the percentage of death and logarithmic value of number of bacteria.

6.2.2. Determination of virulence of clinical isolates of *A. xylosoxidans* in *G. mellonella*

All clinical isolates were prepared as previously described above to produce a final suspension in PBS of $1x10^8$ cells/ml. Larvae were inoculated with 10 µl of bacterial suspension, approximately $1x10^6$ cells, and placed in 24-well plate. Ten worms for each strain, replicated three times. After the injection, larvae were incubated at 37°C and the status of the larvae were assessed every 12 hours until 96 hours after injection. Larvae were considered 'death' when they did not respond to physical stimulation and became melanised (Miyata *et al.*, 2003). Survival of larvae was assessed using Kaplan-Meier survival curves and Log-rank test using the *survival* package in R version 3.1.2.

6.2.3. Virulence genes prediction

To predict potential virulence factors a reciprocal protein BLASTP was performed for each strain predicted proteome against the Virulence Factor Database (VFDB) (Chen, 2012). BLASTP searching was done with a threshold at 1e-5. The top reciprocal BLASTP hits were validated by PSIBLAST against CDD and were curated to ensure consistent results. Resulting genes were assigned back to their OrthoMCL orthologous clusters (Chapter 5) according to gene accession. Finally, OrthoMCL orthologous clusters were transformed into a matrix showing the presence/absence of orthologous groups in each strain.

6.2.4. Phenotype-associated virulence genes prediction

Two sets of genomic data were prepared as inputs for the algorithm. The former was constructed based on VFDB-riched orthologues, and the latter was constructed based on whole genome orthologues. Both inputs for the gene searching were prepared as matrices showing the presence/absence of the orthologues across 25 genomes of tested *A. xylosoxidans*. The rate of kill (fast and slow) was used as a category for Random Forest™ classification. Both matrices were subjected to classification process using Random forest™ in R statistic software. The implementation was previously described in Chapter 4.

6.3. Results

6.3.1. Determination of LD_{50} of type strains, NCIMB 11015 and DSM 2402, in *G. mellonella* infection model

The suitability of *G. mellonella* as an *A. xylosoxidans* infection model was assessed using two type strains. A bacterial titre of 10^6 cells for NCIMB 11015 caused 50% mortality within 24 hours and 100% mortality within 36 hours (Figure 6.1). The same titre for the DSM 2402 strain caused 65 % mortality after 36 hours and 100% mortality within 48 hours (Figure 6.1). The LD_{50} of the two strains shows a significant difference with a mean of 71,969 and 146,757 cells of NCIMB 11015 and DSM 2402, respectively (Student's t-test; $t = -6.0066$; $p < 0.05$) (Figure A4.1). The effect of reducing the inoculum dose was to reduce the rate of larval mortality, and a positive correlation was observed between larval mortality and injected bacterial titre (Figure 6.1). The control larval groups of mock infection and no injection remained alive till the end of the observation period.

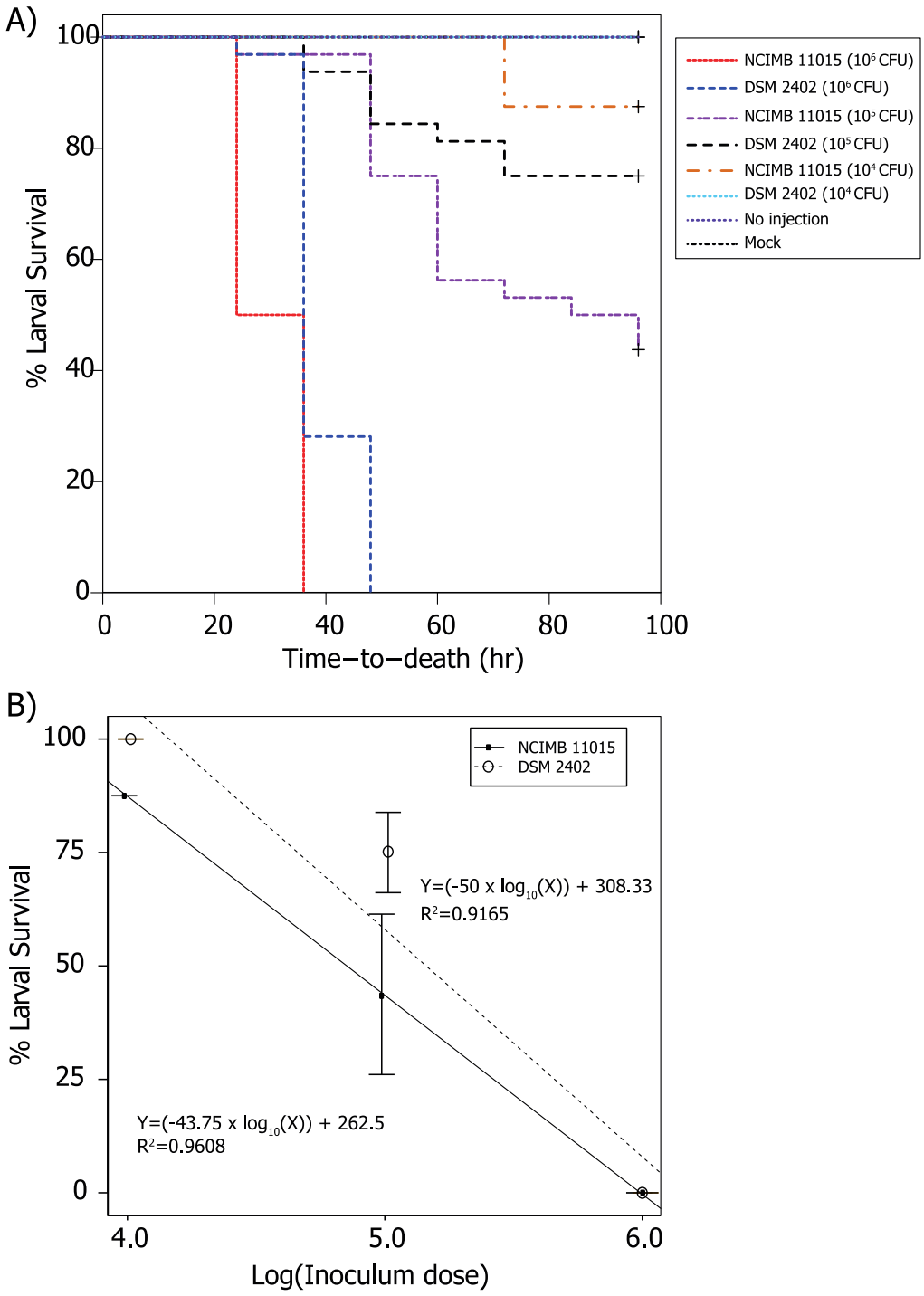


Figure 6.1: Virulence of type strains of *A. xylosoxidans*. A) Kaplan-Meier survival curve of *G. mellonella* after treatment with different concentration of *A. xylosoxidans* strains NCIMB 11015 and DSM 2402. X-axis indicates time after injection. Y-axis indicates the percentage of larval survival. B) A linear regression graph plotted between the percentage of killed larvae and $\log_{10}$(Inoculum dose) from *A. xylosoxidans* strain NCIMB 11015 (black line and black square) and strain DSM 2402 (dash line and open-circle).

6.3.2. Virulence of clinical strains in the Galleria mellonella model

Virulence of clinical strains was determined in *G. mellonella* larvae over a 96 hours post injection. Based on the experiment described in section 6.3.1, it was expected that *A. xylosoxidans* strains would kill at different rates.

Based on the observed differences in Time to Death (TD) at 24 h, the strains were put in two groups (Figure 6.2A). Fast, ≥80% mortality within 24 h and slow <80% mortality within 24 h see Table 6.1. Moreover, the strains were also separated into 2 groups by UPGMA hierarchical clustering (Figure 6.3B).

Table 6.1: Summary table showing groups of strains categorised by time-to-death of *G. mellonella* (Referring to Figure 6.2). NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402. L-strain for Liverpool's clinical isolates; and R-strain for Thai clinical isolates.

Killing group	Strain		
	Type	Liverpool	Thai
Fast ≥80% TD<24 h	NCIMB 11015	L1 L5 L10	R1 R2 R3 R5 R6 R7 R8 R9 R10 R11 R12 R14
Non-fast <80% TD <24h	DSM 2402	L8 L11 L14 L15 L16 L17	R4 R13

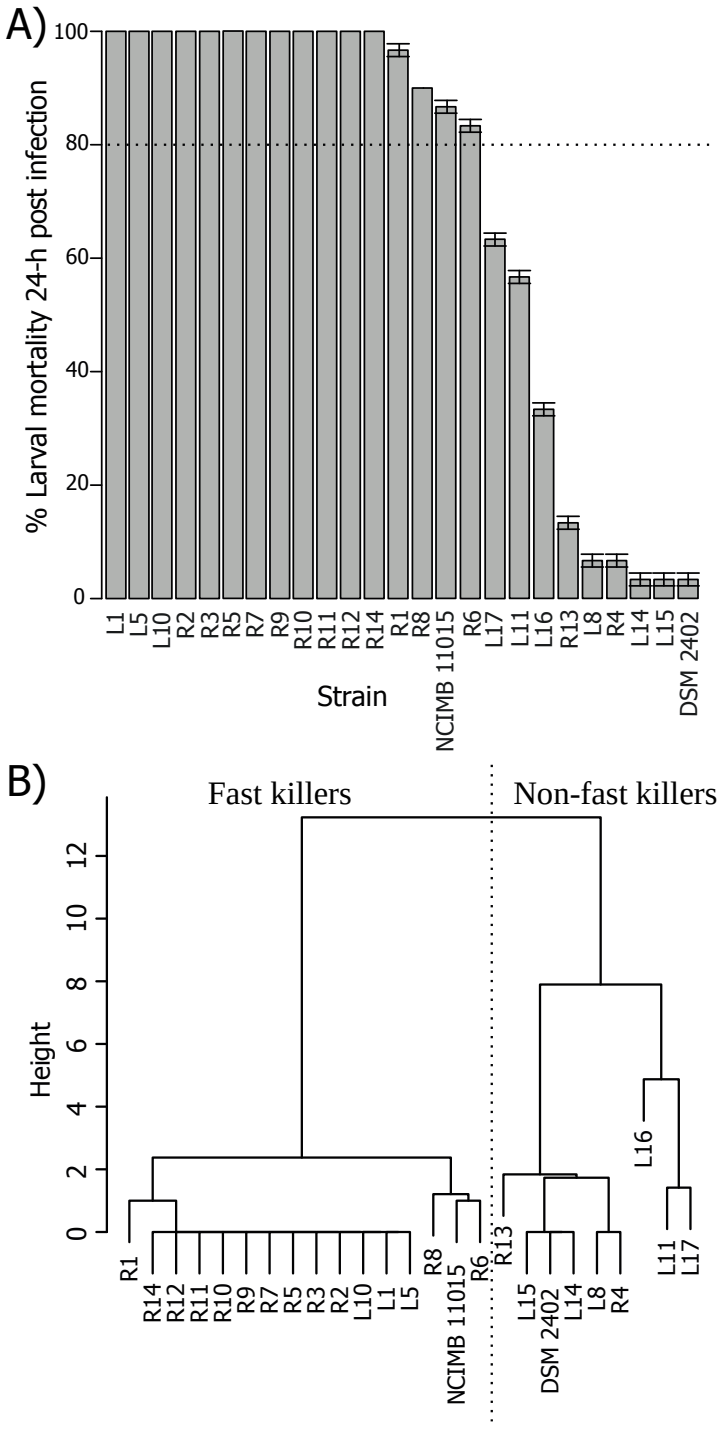


Figure 6.2: Grouping of *A. xylosoxidans* isolates by time-to-death (TD) of the larvae. A) Differential virulence of *A. xylosoxidans* in *G. mellonella* model. Data were collected after 24 hour post-injection B) Hierarchical clustering of the isolates based on TD using UPGMA method, implemented in R. NCIMB 11015, stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; L-strain for Liverpool's clinical isolates; and R-strain for Thai clinical isolates.

6.3.3. Candidate prediction using the VFDB-enriched data set

The VFDB database hits to *A. xylosoxidans* revealed 238 gene clusters containing 4,549 genes (Appendix Table A4.1). Of these orthologous clusters, 76 clusters were shared across all genomes and the most abundant candidates were general secretion pathway protein E, accounting for 50 genes from 25 genomes. Moreover, conserved virulence-associated genes included genes coding flagella biosynthesis, genes coding lipid A and genes involved in capsule biosynthesis (Appendix Table A4.1).

Cluster analysis of the candidate virulence factor genes revealed some local clustering of strains from the same virulence category (Appendix Figure A4.3), for example, fast strain R3, R7 R9, R10; slower strains such as L11, L15 and R13. However, there was little indication of global virulence factor associated with the rate of larval mortality. Interestingly, the heatmap's dendrogram was consistent with core genome phylogeny (Appendix Figure A4.3) and some clinical isolates were genetically related to each other; for example, L11 was related to L15, and L5 was related to R11.

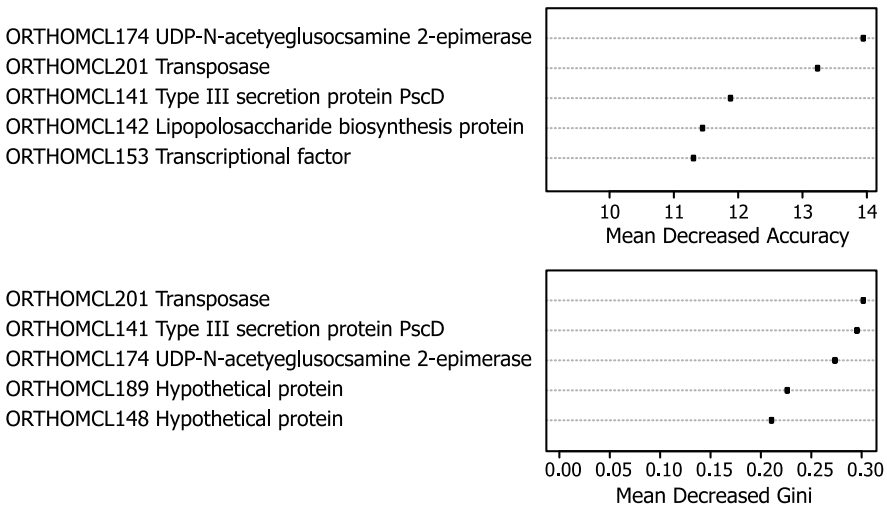


Figure 6.3: A variable importance plot obtained from Random Forest™ classification on virulence of 25 isolates of *A. xylosoxidans* in *G. mellonella* shows the importance of each orthologue based on VFDB-enriched data set. A) Orthologs ranked by their 'mean decreased accuracy' score. B) Orthologous ranked by their 'mean decreased Gini' score. The x-axes indicate the level of importance. The Y-axes show the list of orthologs.

The use of Random Forest™ classification-based algorithm to identify gene candidate that had the highest importance score. The classification extracted three orthologues that had high mean decreased accuracy score and mean decreased Gini score. These orthologues were genes encoding UDP-N-acetylglucosamine-2-epimerase, transposase and type III secretion system protein PscD (As shown in Figure 6.3).

When these genes were considered in the context of the cluster analysis, no strong correlation with rate of larval mortality was observed. However these three genes did correlate with the local clustering discussed above. For example the fast strain cluster R3, R7 R9, R10 all contained type III secretion system protein PscD. It is also evident that most fast strain contained the same PscD gene (Figure 6.4). The PscD was not found on conserved orthologues. UDP-N-acetylglucosamine-2-epimerase was present in most of Fast groups (11/16), and it was not present in any slow strain (Figure 6.4). When investigating gene orthologues of these isolates, the epimerase were found as both conserved genes and acquired genes. Interestingly, none of fast strain contains the transposase, whereas five out of nine of slow strains contain the transposase.

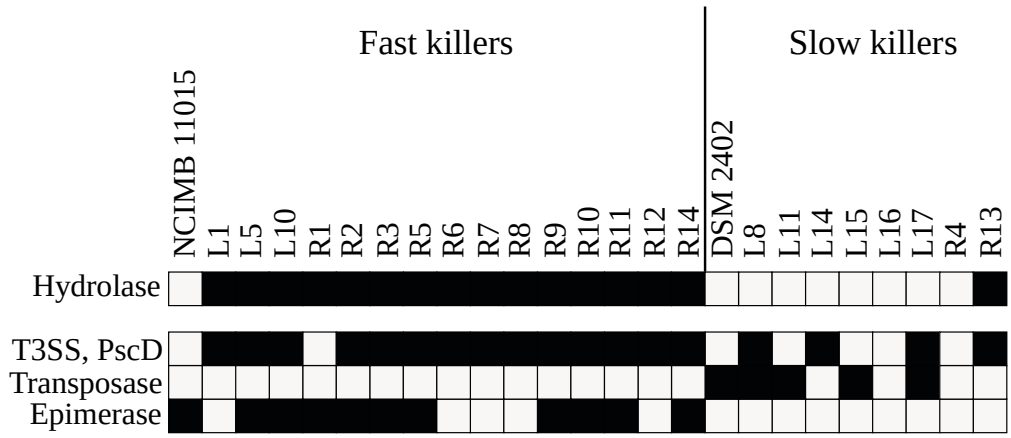


Figure 6.4: The presence of significant Random Forest™ (RF)'s gene groups in *A. xylosoxidans* isolates. Upper panel indicates RF's output from whole genome data set. Lower panels indicate RF's output from VFDB-enriched data set. NCIMB 11015 stands for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; L-strain for Liverpool's clinical isolates; and R-strain for Thai clinical isolates.

6.3.4. Candidate prediction from whole genome data set

The use of Random Forest™ classification-based algorithm to identify gene candidate from the whole genome data sets was able to classify one gene that had correlation with fast larval mortality with a high mean decreased accuracy score and mean decreased Gini score (Figure 6.5). This gene had 80% homology to α/β -hydrolase domain-containing protein from *P. aeruginosa*. This gene was found in 15 (out of 16) fast-killing strains but was also found in one (out of nine) of the slower-killing strains (Figure 6.4). The investigation into the secondary structure of the hydrolase revealed that this protein consisted of α-helices and β-pleated sheet structure (Figure A4.2). The analysis for subcellular localisation of the protein showed that this protein could be cytoplasmic protein.

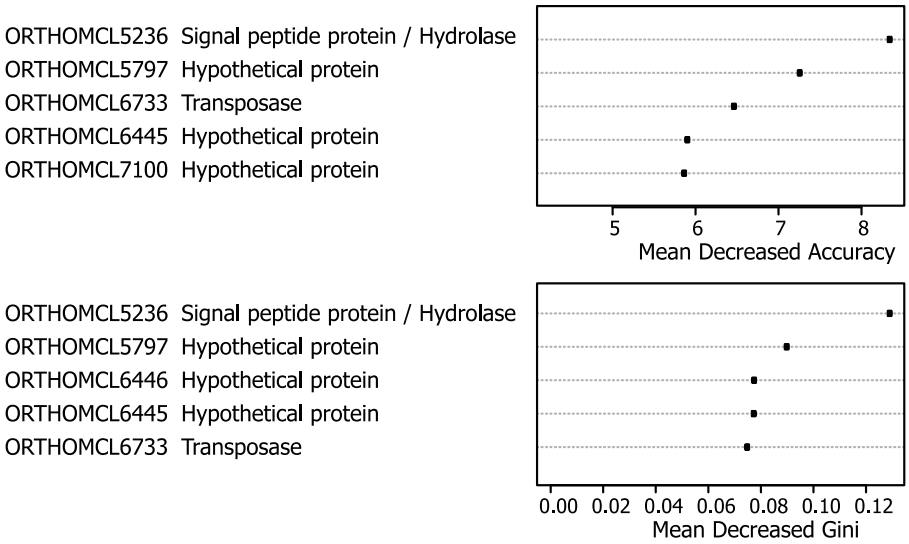


Figure 6.5: A variable importance plot obtained from Random Forest™ classification on the virulence of 25 isolates of *A. xylosoxidans* in *G. mellonella* model shows the importance of each orthologue based on whole genome data set. The X-axes indicate the level of importance. The Y-axes show the list of OrthoMCL's groups.

6.4. Discussion

This is the first study to assess the virulence of *A. xylosoxidans* in *G. mellonella.* Compared to the other animal models for infectious diseases, *G. mellonella* hold many advantages, such as lower in cost. When ethic of animal research considered, using *G. mellnella* attains the 3Rs. Additionally, its immune response is similar to innate immune response in mammals. Nitric oxide production in insect haemocyte is similar to the process in human neutrophil (Bergin *et al.*, 2005). This suggests that *G. mellonella* can be an alternative model to examine *A. xylosoxidans* infection in this study.

6.4.1. *G. mellonella* as an infection model for *A. xylosoxidans*

Preliminary result showed the difference of virulence of *A. xylosoxidans* type strain, as determined by the number of dead insects. The killing of the insects is dose-dependent (Aperis *et al.*, 2007). The experiment demonstrated that *A. xylosoxidans* could establish the infection and could kill the larvae. The larvae could remain undamaged at low inoculum and died at higher inoculum. The lethality of the bacteria was different, as determined by the measurement of LD_{50}. The soil reference strain (NCIMB 11015) was more virulent that the clinical reference strain (DSM 2402) in the insect model. Insects have been used to test the pathogenicity and virulence of human pathogens. *C. elegans* (Ewbank & Zugasti, 2011) and *D. melanogaster* (Dionne & Schneider, 2008) are well-known insect infection models. Insects, including *C. elegans*, *D. melanogaster*, and *G. mellonella*, and mice were used for *P. aeruginosa* infection model and they showed good performance to represent human infection (Tan, Mahajan-Miklos & Ausubel, 1999; Jander, Rahme & Ausubel, 2000; D'Argenio *et al.*, 2001). However, insect models are preferred, when considering 3Rs. The most important advantage of *G. mellonella* over *C. elegans* and *D. melanogaster* is that the great moth larvae can be cultured at 37°C. Taking these observations together, it suggests that *G. mellonella* can be used as an infection model for *A. xylosoxidans*.

6.4.2. Virulence of clinical isolates of *A. xylosoxidans*

Virulence, by definition, is the measurement of the pathogenicity of the bacteria, which, mostly, refers to the ability to cause diseases in the hosts. Using wax moth larvae as a model, injected larvae died of infection whereas non-injected larvae and mock injection larvae did not. This suggests that the bacteria have virulence in the larvae of *G. mellonella*. The difference between the virulence of these strains is probably because of the expression of virulence factors. Gram-negative bacteria posses lipopolysaccharide molecule that induces immune response in innate immune cells that ends up with inflammatory process and septicaemia in human (Mogensen, 2009). However, Montovani *et al.* showed that the virulence of *A. xylosoxidans* was not due to lipopolysaccharide molecule (Mantovani, Levy & Yano, 2012). On account of being a soil bacterium, *A. xylosixidans* can be a natural pathogen for *G. mellonella*. Therefore, a further investigation using a known *G. mellonella* pathogen will be conducted to evaluate the pathogenicity of *A. xylosoxidans*.

As a result of the killing assay, most of Thai isolates exhibited rapid killing by kill more than 80 per cent of the larvae by 24 hours (Figure 6.2). This is similar to the soil type strain. Considering groups of strains by killing rate, the majority of fast-killing strains are strains from Thailand and the majority of slow-killing strains are strains from the UK (Figure 6.2). It can be assumed that strains from different geographical background require different growing conditions such as temperature. Therefore, a further investigation will be conducted by varying incubating temperature of killing assay in *G. mellonella* model.

Subsequently, the hierarchical cluster of virulence genes taken from each strain did not indicate geographical specificity of the strains (Figure A4.3). Furthermore, the cluster revealed groups of fast killers and non-fast killers. As seen in the killing assays, all bacterial isolates are detrimental to the larvae. The factors that cause the variation in time-to-death of the larvae can be different growth rate of the bacteria and different level of virulence factors. Different growth rate of the bacteria can influence time-to-death of the larvae due to difference in the number of bacteria inside the larvae at particular time point. In

other words, bacteria having faster rate of growth can generate a higher number of bacterial cell, when the same amount of time spent, compared to the bacteria having slower growth rate. Subsequently, the fast growing bacteria can cause mortality faster than the slow growing bacteria.

As discussed in Chapter 4 that genome of the species is open, this points to the idea that the difference in virulence of each strain can be due to different virulence genes, which are received from various sources. The evidence of horizontal transferred genes in *A. xylosoxidans* is explained by the detection of mobile element genes, such as integrase (As discussed in Chapter 5), and by the presence of recombination in the species (As discussed in Chapter 4). In other species, lateral gene transfer is considered a tool to exchange virulence genes. For instance, the transfer of Shiga toxin-encoding gene into the *E. coli* via phage, and they become Shiga-toxin-producing *E. coli* strains (Muniesa *et al.*, 2000).

On account of being an emerging and opportunistic pathogen, a few investigations of virulence of *A. xylosoxidans* have been conducted. *A. xylosoxidans* is recognised for its low virulence in healthy individual. This study presented the invasion of *A. xylosoxidans* that ended up with the death of the larvae. The result in this study is in connection with that explained by Hyodo, Katahira & Shigeta (1982) that *A. xylosoxidans* successfully infects and kills immune-compromised mouse. The investigation also detects the distribution of *A. xylosoxidans* throughout the body of mouse, except for the liver. These emphasise the low virulence of *A. xylosoxidans*.

6.4.3. Potential virulence genes in *A. xylosoxidans* based on *G. mellonella* infection model

The virulence factors of *A. xylosoxidans* have not been extensively investigated. (Mantovani, Levy & Yano (2012) detected secretory heat-stable cytotoxic factor from the supernatant of *A. xylosoxidans*. The factor was seen to induce the production of interleukin-6 and interleukin-8, when treating lung cell line. Subsequently, these interleukins caused the inflammation of lung epithelial cells

in *in vitro* investigation. There is only a single study to investigate virulence factor of *A. xylosoxidans.*

There are a number of bacterial virulence-associated genes identified as conserved genes of *A. xylosoxidans*. The most prominent genes are gene coding flagella, capsule, and LPS. These are conventional bacterial virulence deteminants (Casadevall & Pirofski, 2001). However, *A. xylosoxidans* isolates in this study demonstrated variable larval killing ability. The prediction of virulence determinant mediating this differential virulence of *A. xylosoxidans* was therefore conducted.

To identify virulence factor candidates, Random Forest™, together with whole genome data set and VFDB-matched data set, nominated virulence gene candidates. For whole genome dataset, the predictive model ranked hydrolase and hypothetical proteins as statistically virulence-associated genes. For VFDB-enriched dataset, the predictive model ranked UDP-N-acetlyglucosamine 2-epimerase, PscD homologous type 3 secretion system export protein, transposase and hypothetical proteins as statistically virulence-associated genes. Determining biological roles of the candidates, hydrolase, PscD homologous gene, and UDP-N-acetlyglucosamine 2-epimerase were considered potential candidates for the virulence determinants of *A. xylosoxidans* in *G. mellonella* model. The putative functions of hypothetical proteins remain unclear and there is a need for further studies.

6.4.3.1. α/β-hydrolase fold-containing protein

Gene coding α/β-hydrolase fold-containing protein was found as a strong candidate with the highest score assigned by Random Forest™ classifying model. The α/β-hydrolase fold is commonly found as a major part of hydrolytic enzymes, such as esterase, lipase and peroxidase (Nardini & Dijkstra, 1999). This domain serves as a catalyst of several biochemical reactions. A number of bacterial proteins contain this domain have been reported as important factors for homeostasis. For example, *rsbQ* gene encoding α/β-hydrolase domain-containing

protein is essential for energy stress response mechanism in *B. subtilis* (Brody, Vijay & Price, 2001). Another example of α/β-hydrolase-containing protein is 2,4-dioxygenase, which catalyses quinolone degradation in catechol metabolism (Fischer, Künne & Fetzner, 1999). The virulence-associated hydrolase-containing proteins have been reported. For instance, the α/β-hydrolase domain was identified in CFTR inhibitory factor (Cif) from *A. baumannii* (Bahl *et al.*, 2014). Cif acts by inhibiting the expression of CFTR channel to the apical surface of broncho-epithelial layer, so this resembles the characteristic of pulmonary CF (Swiatecka-Urban *et al.*, 2006). A virulence factor BioH from *Francisella*, a causing pathogen of Tularaemia, also contains the hydrolase fold (Feng *et al.*, 2014). These support the potential of α/β-hydrolase-containing protein found in this study to be a virulence factor; however, an actual role of the protein in *A. xylosoxidans* has not been elucidated yet.

6.4.3.2. Type III secretion export protein PscD is a potential virulence protein

For gene candidates obtained from VFDB-enriched data sets, PscD is considered as the most virulence-associated gene. The analysis revealed that PscD-homogous gene is correlated with virulence in the larvae model (Figure 6.4). PscD is a forkhead-associated fold-containing protein and, interestingly, is homologous to inner membrane subunit of type 3 secretion system (T3SS), such as YscD in *Yersinia enterocolitica*, EscD in *E. coli* O157:H7, and CT664 in *Chlamydia trachomatis* (Pallen, Beatson & Bailey, 2005). The T3SS is a membrane-integrated transporter having a needle-like structure (Radics, Königsmaier & Marlovits, 2014). Once, the T3SS is inserted into host cell, an effector molecule is transported through the secretory apparatus directly to the host cell. Previous studies demonstrated the importance of PscD and YscD in virulence of these pathogens. In *Yersinia*, YscD is essential for the secretion of effector protein YopM (Plano & Straley, 1995). In *P. aeruginosa*, PscD is reported to associate with pathogenesis of the infection in great moth model (Miyata *et al.*, 2003) and in zebrafish model (Clatworthy *et al.*, 2009). This observation will probably support the secretory heat-stable protein that causes inflammation in lung epithelial cell line (Mantovani, Levy & Yano, 2012).

Many pathogens use the T3SS to adhere to host cells. For instance, enteropathogenic and enterohaemorrhagic *E. coli* use T3SS to inject Tir protein, which forms a complex with host's cytoskeleton, thus attaching to host cell (Frankel, 2001). Following the attachament to the host cell, the T3SS provides pathogens with a tunnel for virulence protein transfer. The T3SS-dependent virulence proteins damage host cells in two main ways: damaging cytoplasmic process and disrupting cell-to-cell adjacent. The examples of these mechanisms are *P. aeruginosa*'s cytotoxic protein ExoU (Sato *et al.*, 2003) and enteropathogenic *E. coli*'s tight junction-disrupting protein EspG (Tomson *et al.*, 2005). The actual importance and role of PscD protein in *A. xylosoxidans* will be further investigated.

6.4.3.3. Capsule is a potential virulence factor in insect model.

UDP-N-acetlyglucosamine 2-epimerase is an essential enzyme in polysaccharide capsule production. The deletion of this gene in *E. coli* strain K1 causes a problem in capsule biosynthesis (Vann *et al.*, 2004). There were two genes encoding the epimerase on *A. xylosoxidans* genomes. One was included in conserved gene sets and the other was included in accessory gene sets. An extra existence of gene encoding UDP-N-acetlyglucosamine 2-epimerase is also identified in other species. For example, the duplication of UDP-N-acetlyglucosamine 2-epimerase is reported in *S. aureus* with functional redundancy (Kiser *et al.*, 1999).

The polysaccharide capsule is the basic structure found in many pathogenic bacteria, such as *S. pneumoniae* (Kadioglu *et al.*, 2008) and *P. aeruginosa* (Pollack, 1984). The compositions of capsules are different in each species. For example, the capsule of *P. aeruginosa* consists of alginic acid (Jain & Ohman, 2005), while capsular component of *S. pneumoniae* are diverse (Bentley *et al.*, 2006). However, the function of capsule is similar that is to protect bacterial cells from external environment, especially host's immune responses, such as opsonophagocytosis, so as to survive inside the host's body (Wilson, 2002;

Hyams *et al.*, 2010). Furthermore, the presence of capsule makes host recruit more immune cells to the site of infection to eliminate the pathogens, leading to reactive immunopathogenic response, which is an over-inflammatory response to host tissue (Wilson, 2002). Considering the characteristic of virulence, the capsule increases bacterial survival, which subsequently ends up with the inflammation and damage of host. Therefore, the enzyme involved in capsule biosynthesis can be considered as a virulence-associated gene in *A. xylosoxidans*. The mechanism how the capsule manipulates the virulence in *G. mellonella* remains undiscovered.

6.4.3.4. Transposase: An interference of virulence in G. mellonella

This study revealed the possibility of the interruption of bacterial virulence by the insertion of the transposon into the genomes. Mobile genetic element has a significant role in the diversification of bacteria. Transposase is one of the components of the mobile element called 'transposon' (Frost *et al.*, 2005). The mobilisation of transposon follows a pattern of 'cut-and-paste' (Muñoz-López & García-Pérez, 2010) and the mobilisation can be promoted by transformation (Domingues *et al.*, 2012), refering to an open pan-genome characteristic of the isolates (As discussed in chapter 4). It is assumed that the presence of the obtained transposase gene in low-virulence strain can be aasociated with the elimination of virulence gene by the insertion of transposon. However, this can happen by chance. Therefore, an additional study to investigate the importance of transposase gene in the virulence of *A. xylosoxidans* infection should be conducted.

6.5. Conclusion and Future work

The findings of this study demonstrate the integration of laboratory-based experiments and bioinformatic analysis in order to investigate virulence of emerging pathogens. The study displayed that *G. mellonella* is suitable for a model to study *A. xylosoxidans* infection. The model is convenient and easy to handle with due to its inexpensive cost. The larvae model can also be a replacement for mammalian model, when reviewing the principles of 3Rs.

The finding presented herein reveals the difference in virulence of *A. xylosxidans* with the use of *G. mellonella*. The investigation shows that environmental strain, NCIMB 11015, is more virulent than clinical strain, DSM 2402. However, some Thai clinical isolates demonstrate aggressive larvae killing. The combination of bioinformatic approaches, whole genome sequence and the killing of larvae has identified α/β-hydrolase-containing protein from whole genome data set, and UPD-N-acetylglucosamin-2-epimerase and T3SS protein PscD from virulence gene-enriched data. The best candidate is α/β-hydrolase-containing on account of its biological function and its existence in the isolates.

The limitations of this study are that the number of bacteria was not determined after the larval infection and the distribution of bacteria inside the larvae was not conducted. Moreover, the assay at different temperatures has been performed yet. The cellular mechanism of *A. xylosoxidans* in *G. mellonella* model is still questioned. Nevertheless, The presence of these genes encourages the eagerness to conduct in-depth study of those candidates and the virulence using more comprehensive methods such as single molecule sequencing to determine the presence of transposon on the genome and mutagenesis of α/β-hydrolase-containing protein, epimerase and PscD to observe the virulence in infection model.

Chapter 7

General discussion

This study attempted to apply a wide range of methods in biology to understand *A. xylosoxidans* on multiple levels – species identification, pan-genome analysis, antibiotic susceptibility, and virulence. This study also sought to explore the functional and comparative genomic perspective of the antibiotic resistance and virulence of *A. xylosoxidans*. As *A. xylosoxidans* has been increasingly identified from clinical sites, CF centres in particular, and has been perceived as an opportunistic pathogen, investigating genomic attribute, as well as a factor that is possibly responsible for antibiotic resistance or virulence of the pathogen, may allow relevant authorities to establish an appropriate treatment and infection control, especially in the hospital. This study led to the discussion of these following topics:

1. Is *A. xylosoxidans* an emerging pathogen?
2. How does the study of microbial pan-genome explain the events of hospital-acquired infections and antibiotic resistance in bacteria ?
3. Do whole genome sequencing and bioinformatics help to understand emerging pathogens?

7.1. Is *A. xylosoxidans* an emerging pathogen?

The empirical data to answer this question are summarised by the demonstration of phylogenetic relationships between isolates. The phylogenetic relationships constructed using different sets of genes, including core genes set, antibiotic resistance-associated genes and virulence-associated genes, revealed the diverse population structure of *A. xylosoxidans*, as stated in Chapter 3, 4, 5, and 6. The genome analyses demonstrated no significant differentiation between clinical isolates and an environmental isolate. Moreover, most of the clinical isolates in this study were collected from patients with underlying conditions, for example,

CF, COPD, and cardiovascular and metabolic diseases. Patients with these diseases are likely to be susceptible to infections.

Emerging pathogens, by definition, are microorganisms that can colonise and cause diseases to hosts with these three possible mechanisms: new hosts, new phenotypes and new area (Engering, Hogerwerf & Slingenbergh, 2013). Opportunistic pathogens, organisms that can cause infection, takes advantage over abnormal conditions of host such as chronic diseases, persistent infection by other pathogen, and ageing (Brown, Cornforth & Mideo, 2012). Previous studies revealed that most of the cases with *A. xylosoxidans* infection have underlying diseases (Kanellopoulou *et al.*, 2004; Rønne Hansen *et al.*, 2006; De Baets *et al.*, 2007; Amoureux *et al.*, 2012; Mantovani, Levy & Yano, 2012; Trancassini *et al.*, 2014). The collection of isolates in this study supported the theories of emerging and opportunistic pathogen as those clinical isolates were collected from patients with chronic diseases and from hospitalised patients. The patients with underlying diseases are prone to be susceptible to infection, so they can be infected with bacteria that inhabit the environment. Following the hospitalisation, patients can bring those environmental bacteria to the hospital, which allows the circulation of the bacteria in the hospital. Those bacteria can cause serious hospital-acquired infection if they are resistant to antibiotics.

As an emerging pathogen, the availability of robust methods for *A. xylosoxidans* identification is restricted. An accurate identification of the pathogens of hospital-acquired infection is required in order to give an appropriate treatment for the infection. There is a use of multiple approaches for the species identification of emerging pathogens (Dong *et al.*, 2008; Emerson *et al.*, 2008). Focusing merely on *A. xylosoxidans*, several studies have suggested various approaches such as 16S rDNA-directed PCR (Liu *et al.*, 2002) and MLST (Spilker, Vandamme & LiPuma, 2012). This study demonstrated the limitation of single identification method to identify *A. xylosoxidans*. The analysis, therefore, proposed that the hierarchical combination of identification would be a suitable approach for *A. xylosoxidans* identification. This strongly suggests that multiple identification methods are essential for the identification of bacteria, especially for the uncommon and emerging pathogenic bacteria.

Despite the fact that using multiple identification approaches may improve the accuracy of species identification for emerging pathogens, facilities for the identification, such as a sequencer and MALDI-TOF, are unavailable in community hospitals, especially hospitals in developing countries including Thailand. Phenotypic test is widely used in those hospitals and the test provides healthcare providers with reliable results for well-known human pathogens, not for emerging pathogens. Therefore, patients infected by emerging pathogens have to be transferred to research-led hospitals, where advanced facilities are available. A precise identification is essential for the epidemiological study of the pathogens; however, giving an appropriate treatment is more clinically important. This makes antibiotic susceptible profile and the identification of antibiotic resistance genes become more important than species identification in clinical practice. Hence, antibiotic susceptible test should be available and performed in all hospitals, primary unit in particular, to shape the administration of an appropriate treatment for the infection. The species identification, which is second important, can be performed afterward by hospitals or institutes having the advanced facilities. This strategy may result in an improvement in treatment and epidemiology of the infections.

7.2. How does the study of microbial pan-genome explain the events of hospital-acquired infections and antibiotic resistance in bacteria?

The analyses of pan-genome and antibiotic resistance of *A. xylosoxidans* enabled us to understand the genomic characteristics that have major impacts on hospital-associated infection. One of the important problems of hospital-associated pathogens is antibiotic resistance (Struelens, 1998), specifically in Gram-negative bacteria (Lynch, Clark & Zhanel, 2013). The antibiotic resistance of *A. xylosoxidans* has been recognised as intrinsic phenotypes (Bador *et al.*, 2011, 2013; Hu *et al.*, 2015). Similar to previous studies (Bador *et al.*, 2011, 2013), this study revealed natural antibiotic resistance to aminoglycosides, fluroquinolones and cephalosporins, in association with the presence of RND-type efflux pumps.

The analyses of genomic characteristic also revealed the capability of *A. xylosoxidans* to obtain antibiotic resistance genes from external sources, as explained by the open pan-genome and evidenced by the presence of integron. Bacteria with open-genome have shown dynamic property of the genomes (Medini *et al.*, 2005). The open pan-genome also refers to continuous exchange of genes between bacteria in order to survive in new environment. Regarding hospital-associated infections, bacteria with open pan-genome carry a prominent feature that can transfer antibiotic resistance genes and obtain virulence genes from the environment. Open pan-genome characteristic can be seen in several pathogenic bacteria, including *S. pneumoniae*, *S. agalactiae*, *E. coli*, and *E. faecium*. Common features determined by the opened genome of those pathogens are the ability to exchange their genetic with environements, as demonstrated by the retrieval of antibiotic resistance genes in *S. pneumoniae* and the retrieval of virulence genes in *E. coli*. On the other hands, close pan-genome can be observed in some pathogenic bacteria, such as *Bordetella* species. The core-genome of *Bordetella* contributed to >95% of the genome, in which virulence genes are already included. The high conservation of *Bordetella* genome can be a consequence of sporulation, providing preventive shells for the genome from harsh environments.

The population structure of the species reveals the diversification of the isolates, as presented by no geography-specific isolates based on the analysis of either core genome or accessory genome. A similar population structure, which is diverse, was seen in the genomic analysis of *Haemophilus parasuis*, demonstrated by mixed MLST and BAPS population structure analysis (Howell, 2014). However, the population structure of *A. xylosoxidans* in this study does not clearly represent the actual population structure of the species because the number of isolates/genomes included may not reach the level that geographical specificity can be discriminated.

A. xylosoxidans is also well-known for the potential to obtain additional resistance via horizontal gene transfers (Traglia *et al.*, 2012; Di Pilato, Pollini & Rossolini, 2014; Chen *et al.*, 2014; Wittmann *et al.*, 2014). The existence of lateral genetic transfers may lead to unusual and harmful features: antibiotic

resistance and virulence (Ochman, Lawrence & Groisman, 2000). Under the pressure of antibiotic uses in the hospistals, genetic elements help bacteria to survive, by providing antibiotic resistance genes (Rowe-Magnus & Mazel, 2002). The results in this study strengthened those findings by presenting the insertion of resistance-associated integron in highly resistant *A. xylosoxidans* isolates. This study also displayed the transfer of mobile genetic elements in the hospitals, as exemplified by the presence of identical integron in *A. baumannii* and in different strain-types of *A. xylosoxidans*. In some bacteria, the acquisition of external genetic elements is also important for the transformation from non-virulent strains to virulent strains; for instance, the transformation of *E. coli* into Shiga toxin-producing *E. coli* by obtaining Stx1 gene-containing phages (Muniesa *et al.*, 2000). The procurement of prophages that is associated with the enhanced competitiveness of the clone (Winstanley *et al.*, 2009) also exemplifies the importance of lateral gene transfer to induce virulence in bacteria.

The integration of genetic elements into bacterial genomes leads to genetic recombination, allowing for molecular evolution of bacteria (Smets & Barkay, 2005). The recombination accelerates the evolution of the microorganisms faster than a single-nucleotide mutation due to the replacement of a large piece of DNA sequence, rather than a single nucleotide of an existing gene(Didelot *et al.*, 2012; Boto, 2010). Recombination has a significant impact upon the population's structure of some pathogenic bacteria, including *P. aeruginosa* (Darch *et al.*, 2015), *E. faecum* (de Been *et al.*, 2013), and *S. aureus* (Feil *et al.*, 2003). In addition, the recombination contributes to the outbreak of *L. pneumophila* (Sánchez-Busó *et al.*, 2014) and *S. aureus* (Castillo-Ramírez *et al.*, 2011, 2012). Recently, genomic analysis has shown the low rate of recombination in genus *Achromobacter*. Inversely, the analysis presented herein the significant level of recombination on the core genomes, as evaluated by PHI test, BratNextGen and the non-congruent phylogenetic tree. The importance of bacterial recombination for hospital-associated infection is that the recombination drives the diversity of bacteria (Boinett & Cain, 2014). Therefore, it enhances the probability of the pathogens to survive since there is the diverse population of bacteria. If the events involve genes, which are detrimental to the host, including antibiotic resistance genes and genes coding virulence determinants, the recombined

bacteria will be more harmful, leading to serious infection in the hospital. In other words, the species that are naturally recombinant are likely to be a reservoir for those detrimental genes.

Considering the diversity of the species, *A. xylosoxidans* seems to be a broad diverse species as there was no particular pathogenic sequence type. Inversely, other pathogens, such *E. coli* (Alghoribi *et al.*, 2014) and *E. faecium* (de Been *et al.*, 2013), have specific pathogenic sequence types or clonal. Importantly, with the open pan-genome of *A. xylosoxidans*, the pathogen can cause problems in hospital-acquired infections as it can habour and transfer virulence- or drug resitance-associated mobile genetic elements to other bacteria.

Another principal contribution of this study has been moved forward to understand the antibiotic resistance problem. Since a hospital is recognised as a place where various antibiotic agents are extensively used, the inappropriate use of the antibiotic can induce the adaptation of bacteria - from susceptible strains to resistant strains. Importantly, empirical medication is often considered to treat infections. For example, aminoglycosides are usually used as empirical medication for Gram-negative infections. The investigation into the comparison between the antibiotic resistance susceptibility of clinical isolates of *A. xylosoxidans* and the environmental isolate of *A. xylosoxidans* revealed that Thai isolates were more resistant to antibiotics than British isolates. Considering the regulation of antibiotic use, medical prescription is always required to purchase antibiotics in the U.K. None of them is over-the-counter medicine. Therefore, the distribution of antibiotic is under inspection that reduces an inappropriate use of the agents. On the other hand, the regulation is poor-controlled in Thailand. The first possible contributor to this problem is the regulation for antibiotic prescription in Thailand. Some first-line antibiotics, such as amoxicillin, are over-the-counter medicines. They can be purchased without prescription from anywhere from chemists to local groceries. The second contributor is the medication compliance of Thai people. Most of Thai people stop taking medication when they feel that the symptoms are relieved (Personal communication with Dr. Ponpan Matangkasombut). The failure to complete the recommended course of antibiotic leads to the development of resistance strains

of bacteria. Antibiotic resistant bacteria, as a consequence of the failure, can be introduced to the hospital when those patients are hospitalised. It would seem sensible to determine the possession of antibiotic resistance associated genes in the communities, including hospitals and residential area, in order to investigate the evolution and the transmission of those genes.

7.3. Does whole genome sequencing and bioinformatics help to understand emerging pathogens?

Sequencing technologies and bioinformatics have become essential tools for a study of molecular microbiology. The improvement of facilities for whole genome sequencing has allowed for an increasing number of genome sequenced. In the latest publicly available Genbank information (September, 2015), there are over 20 million bacterial sequences, including uncultured bacteria (Figure 7.1). *E. coli* contributes to the most abundant bacterial nucleotide sequences (1,406,795 deposited sequences, 6.95%). Clinically relevant species are the targets of sequencing, as illustrated by the top ten most sequenced species which are the causes of infection in human (Figure 7.1). Nucleotide sequencing has also been used as an unlocking tool for the investigation of emerging pathogens, such as *M. abscessus* (Choo *et al.*, 2014), *H. influenzae* type f (Su *et al.*, 2014), *Vibrio furnissii* (Lux, Lee & Love, 2014), and nosocomial pathogens including *S. aureus* (Harris *et al.*, 2010; Price *et al.*, 2014).

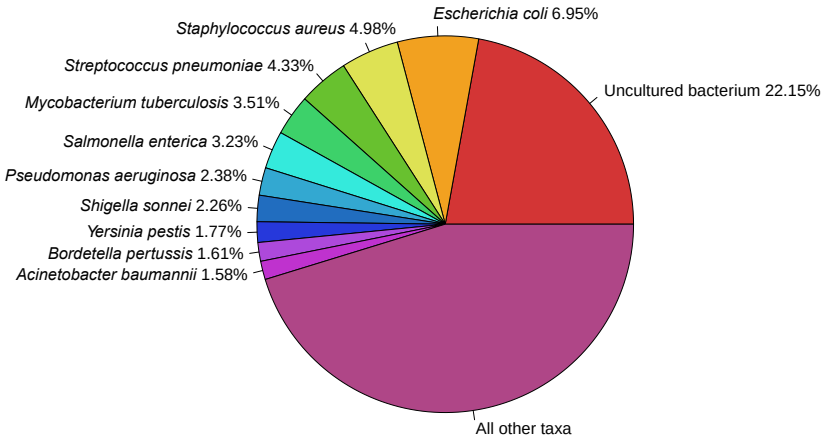


Figure 7.1: Pie chart of bacterial sequences available on Genbank database in September, 2015.

This study applied whole genome sequencing to build up the understanding of *A. xylosoxidans*. The sequencing was, initially, performed on Illumina HiSeq platform, in which generated 150-bp paired-end reads. The critical step in genome sequencing approach is to obtain the best assembly, *de novo* assembly in particular. Statistical parameters have been used to evaluate the efficiency of the assemblers. N50 values are widely used to determine the quality of the assemblies. However, the investigation into the genome with the highest N50 is required to determine the quality of the assemblies in the event of repetitive sequences the can be ignored by the assemblers (Baker, 2012). The comparison of the efficiency of assembly between Velvet and Spades was performed. Spades assembler is likely to have a better ability (Bankevich *et al.*, 2012), as determined by N50 values, the number of contigs and the size of the largest contigs. SMRT sequencing using PacBio platform has been introduced to achieve a single contig from the genome sequencing. The combination of Illumina's sequencing and PacBio's sequencing allows for complete genome sequences of bacteria such as *B. subtilis* (Kamada *et al.*, 2014). Likewise, the sequencing using PacBio RS II platform, as well as Illumina platform, also revealed complete genome sequences of *A. xylosoxidans*. With a single contig, the localisation of antibiotic resistance-carrying integron on the genomes of antibiotic resistance isolates R4 and R8 can be achieved.

In parallel with sequencing technologies, the number of software and database for genomic analyses has also been increasing. Apart from Genbank database (http://www.ncbi.nlm.nih.gov/genbank), several databases are established to provide information for the study of specific genus or species of bacteria, for example, the *Pseudomonas* Genome DB for *Pseudomonas* (http://www.pseudomonas.com) and EcoCyc for *E. coli* K-12 MG 1655 (http://ecocyc.org). There are also databases for specific study such as PHAST for bacteriophage identification (http://phast.wishartlab.com) and PubMLST for MLST (http://pubmlst.org). For pathogenic bacteria specifically, the number of databases has enabled the comparative analysis of antibiotic resistance-associated genes and virulence-associated genes. However, emerging pathogens and uncommon pathogens are not included in some databases such as VFDB (Chen, 2004). Therefore, the prediction of virulence genes of those uncommon

pathogens, such as *Erysipelothrix rhusiopathiae* (Kwok *et al.*, 2014), is mostly performed by means of alternative approach based on sequence homology. Identifying virulence genes and antibiotic genes in emerging pathogens remains a main challenge.

Data mining is vital for genomic analysis since whole genome sequence contains large size of information. The critical approach for genomic analysis is to explore the correlation and association between genetic information and biological information of the organisms. Several classification and correlation analyses have been developed and applied to examine genotype-phenotype correlation. This investigation executed unsupervised classifying approach, using Random Forest™, to find the association between phenotypes and whole genome data set. Random Forest™ algorithm has been implemented in several 'omic' studies. As Random Forest™ can deal with large data set, it has been used as an approach to reduce the number of candidates for further analysis; for instance, the number of psychiatric disease-associated SNPs was narrowed by Random Forest™ (Lunetta *et al.*, 2004). Nevertheless, with high-dimensional data set, the ability of Random Forest™ decreases (Winham *et al.*, 2012). This indicates that filtered data set is more suitable for Random Forest™. It is, therefore, noteworthy that results obtained from Random Forest™ are 'statistically' important, and they are not always biologically important in the observed variables. The analysis for antibiotic resistance-associated genes demonstrated that Random Forest™ analysis resulted in genes that have been proved to play a role in antibiotic resistance. However, there is no obvious evidence to show that genes are statistically associated with virulence in *G. mellonella* model. Insight studies of the role of those genes will be necessary to pursue as a future research.

7.4. Future work

The study of the virulence of *A. xylosoxidans* identified genes important for the pathogenicity of the species using *G. mellonella* infection model. However, the assessment of the virulence of the candidates in *in vivo* experiments is still required. For further study, bacteria with knocked-out virulence gene candidates

will be tested using *G. mellonella* model to investigate whether those genes contribute to the pathogenicity of the bacterial. In parallel, the distribution of the pathogen will be examined using green fluorescent-tagged bacteria. Additional analysis can be conducted further in mouse model. Hyodo, Katahira & Shigeta (1982) demonstrated that adaptive immunity responsible for *A. xylosoxidans* infection in BALB/c mice. Further investigation can be conducted in CF mouse model using F508del-CFTR mice (Wilke *et al.*, 2011). Considering the principles of 3Rs, investigating chronic infection in CF mouse model cannot be replaced by other non-mammalian model. The investigation in CF mouse model will be executed using signature-tagged mutagenesis as the approach can identify genes essential for the pathogenicity of the bacteria in tested animal model. Then the features of the virulence gene candidates will be studied in terms of roles in metabolic pathways, biological effects on hosts, and three-dimentional structure of the protein. These studies will pave the way for specific treatments, such as species-specific monoclonal antibody and passive immunisation.

As the pathogen is well known for its multidrug resistance, a rapid detection for antibiotic resistance in pathogenic bacteria will improve the precision of antibiotic administration, without commencing treatment with non-specific empirical antibiotic. Therefore, the rapid identification for resistance genes will be developed to shorten time-to-result of currently used antibiogram. In addition, an inhibitor for RND-type efflux pumps will be invented as the activity of the RND-type efflux pumps play an important role in antibiotic resistance in severl human pathogenic bacteria, including *A. xylosoxidans*, *A. baumannii*, and *P. aeruginosa*.

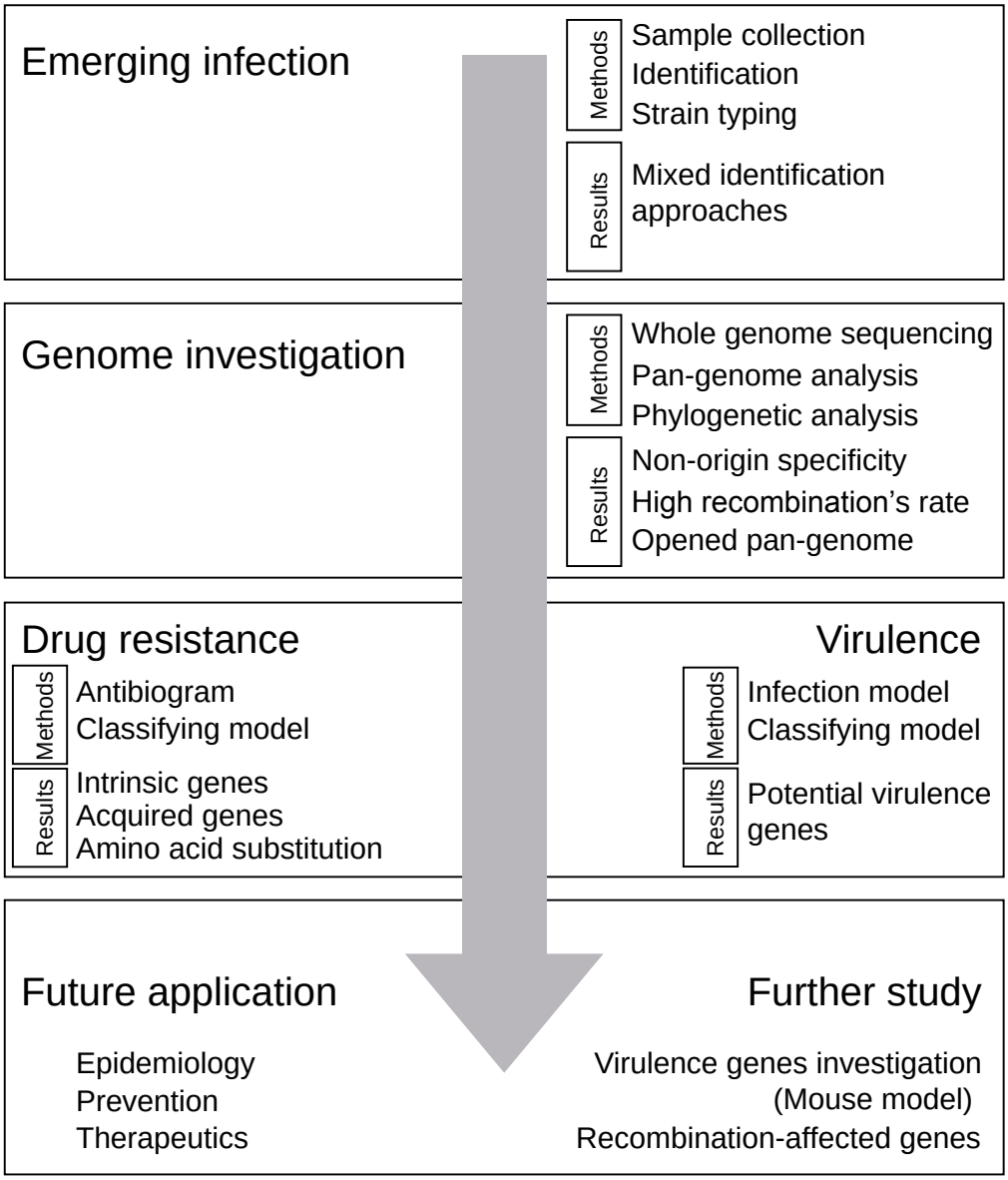


Figure 7.2: A study flow diagram for the investigation of *A. xylosoxidans*

7.5. Conclusion

This study provides the comprehensive analysis of an emerging pathogen, *A. xylosoxidans*, using advanced genome sequencing technologies and bioinformatics (Figure 7.2). The study has also shown the limitation of (1) species identification for this emerging pathogen, and (2) the comparative genomic analysis with regards to antibiotic resistance and virulence in larval model. The scale of this investigation is applicable to the explanation for *A. xylosoxidans* in the local area. A large number of experimental observations and analyses are required to understand the event of *A. xylosoxidans* infection in a

broad level. There is, therefore, a need for larger collections of isolates and more various origins of isolates. The genomic study of other emerging nosocomial pathogens will offer a comprehensive situation of hospital-acquired infection, providing molecular evidences for public health policy maker to prevent an outbreak as well as antibiotic resistance in hospital setting. Finally, an animal infection model is required for the investigation of virulence.

References

AbdulWahab, A. *et al.* (2015) Discrepancy in MALDI-TOF MS identification of uncommon gram-negative bacteria from lower respiratory secretions in patients with cystic fibrosis. *Infection and Drug Resistance*. 883–88.

Abraham, E.P. & Chain, E. An enzyme from bacteria able to destroy penicillin. 1940. *Reviews of infectious diseases*. [Online] 10 (4), 677–678. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3055168 [Accessed: 14 February 2015].

Adam, P. *et al.* (2013) Prevalence of Achromobacter xylosoxidans in pulmonary mucosa-associated lymphoid tissue lymphoma in different regions of Europe. *British journal of haematology*. [Online] Available from: doi:10.1111/bjh.12703.

Adams, E. (1967) The antibacterial action of crystal violet*. *Journal of Pharmacy and Pharmacology*. [Online] 19 (12), 821–826. Available from: doi:10.1111/j.2042-7158.1967.tb09550.x [Accessed: 21 September 2015].

Alavi, P. *et al.* (2014) Stenotrophomonas comparative genomics reveals genes and functions that differentiate beneficial and pathogenic bacteria. *BMC genomics*. [Online] 15 (1), 482. Available from: doi:10.1186/1471-2164-15-482 [Accessed: 28 July 2014].

Aldred, K.J., Kerns, R.J. & Osheroff, N. (2014) Mechanism of quinolone action and resistance. *Biochemistry*. [Online] 53 (10), 1565–1574. Available from: doi:10.1021/bi5000564 [Accessed: 18 February 2015].

Alghoribi, M.F. *et al.* (2014) Galleria mellonella Infection Model Demonstrates High Lethality of ST69 and ST127 Uropathogenic E. coli. Dipshikha Chakravortty (ed.). *PloS one*. [Online] 9 (7), e101547. Available from: doi:10.1371/journal.pone.0101547 [Accessed: 5 September 2014].

Ali, A. *et al.* (2015) Pan-genome analysis of human gastric pathogen H. pylori: comparative genomics and pathogenomics approaches to identify regions associated with pathogenicity and prediction of potential core therapeutic targets. *BioMed research international*. [Online] 2015139580. Available from: doi:10.1155/2015/139580 [Accessed: 26 August 2015].

AlonsoDeVelasco, E. *et al.* (1995) Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines. *Microbiological reviews*. [Online] 59 (4), 591–603. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239389&tool=pmcentrez&rendertype=abstract [Accessed: 26 March 2015].

Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *Journal of Molecular Biology*. [Online] 215 (3), 403–410. Available from: doi:10.1016/S0022-2836(05)80360-2.

Amoureux, L. *et al.* (2013) Detection of Achromobacter xylosoxidans in hospital, domestic, and outdoor environmental samples and comparison with human clinical isolates. *Applied and environmental microbiology*. [Online] Available from: doi:10.1128/AEM.02293-13.

Amoureux, L. *et al.* (2012) Epidemiology and resistance of Achromobacter xylosoxidans from cystic fibrosis patients in Dijon, Burgundy: First French data. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] Available from: doi:10.1016/j.jcf.2012.08.005.

Anderson, A.C. *et al.* (2014) Rapid species-level identification of vaginal and oral lactobacilli using MALDI-TOF MS analysis and 16S rDNA sequencing. *BMC microbiology*. [Online] 14 (1),

312. Available from: doi:10.1186/s12866-014-0312-5 [Accessed: 27 May 2015].

Anderson, B.E. *et al.* (1991) Ehrlichia chaffeensis, a new species associated with human ehrlichiosis. *Journal of Clinical Microbiology*. [Online]. 29 (12) p.2838. Available from: /pmc/articles/PMC270443/?report=abstract [Accessed: 7 September 2015].

Antunes, P. *et al.* (2005) Dissemination of Sulfonamide Resistance Genes (sul1, sul2, and sul3) in Portuguese Salmonella enterica Strains and Relation with Integrons. *Antimicrobial Agents and Chemotherapy*. [Online] 49 (2), 836–839. Available from: doi:10.1128/AAC.49.2.836-839.2005 [Accessed: 5 April 2015].

Aperis, G. *et al.* (2007) Galleria mellonella as a model host to study infection by the Francisella tularensis live vaccine strain. *Microbes and infection / Institut Pasteur*. [Online] 9 (6), 729–734. Available from: doi:10.1016/j.micinf.2007.02.016.

Apidianakis, Y. & Rahme, L.G. (2011) Drosophila melanogaster as a model for human intestinal infection and pathology. *Disease models & mechanisms*. [Online] 4 (1), 21–30. Available from: doi:10.1242/dmm.003970 [Accessed: 27 April 2015].

Apisarnthanarak, A. *et al.* (2006) Inappropriate antibiotic use in a tertiary care center in Thailand: an incidence study and review of experience in Thailand. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*. [Online] 27 (4), 416–420. Available from: doi:10.1086/503348.

Apisarnthanarak, A. *et al.* (2008) Nonjudicious dispensing of antibiotics by drug stores in Pratumthani, Thailand. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*. [Online] 29 (6), 572–575. Available from: doi:10.1086/587496.

Apisarnthanarak, A. & Mundy, L.M. (2008) Correlation of antibiotic use and antimicrobial resistance in Pratumthani, Thailand, 2000 to 2006. *American journal of infection control*. [Online] 36 (9), 681–682. Available from: doi:10.1016/j.ajic.2007.10.022.

Arroyo, J.C. *et al.* (1987) Diversity of plasmids in Achromobacter xylosoxidans isolates responsible for a seemingly common-source nosocomial outbreak. *Journal of clinical microbiology*. [Online] 25 (10), 1952–1955. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=3667915&retmode=ref&cmd=prlinks.

Ashiru-Oredope, D. & Hopkins, S. (2013) Antimicrobial stewardship: English Surveillance Programme for Antimicrobial Utilization and Resistance (ESPAUR). *The Journal of antimicrobial chemotherapy*. [Online] 68 (11), 2421–2423. Available from: doi:10.1093/jac/dkt363 [Accessed: 27 July 2015].

Atalay, S. *et al.* (2012) Clinical and microbiological evaluation of eight patients with isolated Achromobacter xylosoxidans. *Scandinavian journal of infectious diseases*. [Online] Available from: doi:10.3109/00365548.2012.664780.

Auckland, C. *et al.* (2002) Linezolid-resistant enterococci: report of the first isolates in the United Kingdom. *The Journal of antimicrobial chemotherapy*. [Online] 50 (5), 743–746. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12407134 [Accessed: 19 June 2015].

Bador, J. *et al.* (2011) First description of an RND-type multidrug efflux pump in Achromobacter xylosoxidans, AxyABM. *Antimicrobial agents and chemotherapy*. [Online] 55 (10), 4912–4914. Available from: doi:10.1128/AAC.00341-11.

Bador, J. *et al.* (2013) Innate Aminoglycoside Resistance of Achromobacter xylosoxidans Is Due

to AxyXY-OprZ, an RND-Type Multidrug Efflux Pump. *Antimicrobial agents and chemotherapy*. [Online] 57 (1), 603–605. Available from: doi:10.1128/AAC.01243-12.

De Baets, F. *et al.* (2007) Achromobacter xylosoxidans in cystic fibrosis: prevalence and clinical relevance. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] 6 (1), 75–78. Available from: doi:10.1016/j.jcf.2006.05.011 [Accessed: 26 August 2014].

De Baets, F. *et al.* (2013) Achromobacter xylosoxidans induced bronchiolitis obliterans in cystic fibrosis. *Pediatric pulmonology*. [Online] Available from: doi:10.1002/ppul.22864.

Bahl, C.D. *et al.* (2014) Signature motifs identify an Acinetobacter Cif virulence factor with epoxide hydrolase activity. *The Journal of biological chemistry*. [Online] 289 (11), 7460–7469. Available from: doi:10.1074/jbc.M113.518092 [Accessed: 14 August 2015].

Baker, D. *et al.* (2011) Critical appraisal of animal models of multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England)*. [Online] 17 (6), 647–657. Available from: doi:10.1177/1352458511398885 [Accessed: 10 August 2015].

Baker, M. (2012) De novo genome assembly: what every biologist should know. *Nature Methods*. [Online] 9 (4), 333–337. Available from: doi:10.1038/nmeth.1935 [Accessed: 17 December 2014].

Bankevich, A. *et al.* (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. [Online] 19 (5), 455–477. Available from: doi:10.1089/cmb.2012.0021.

Barber, M. (1961) Methicillin-resistant staphylococci. *Journal of Clinical Pathology*. [Online] 14 (4), 385–393. Available from: doi:10.1136/jcp.14.4.385 [Accessed: 19 June 2015].

Barford, J. & Coates, A. (2009) The pathogenesis of catheter-associated urinary tract infection. *Journal of Infection Prevention*. [Online] 10 (2), 50–56. Available from: doi:10.1177/1757177408098265 [Accessed: 15 June 2015].

Barraud, O. *et al.* (2013) An antibiotic-resistant class 3 integron in an Enterobacter cloacae isolate from hospital effluent. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. [Online] 19 (7), E306–E308. Available from: doi:10.1111/1469-0691.12186 [Accessed: 28 July 2015].

Bayjanov, J.R. *et al.* (2012) PhenoLink - a web-tool for linking phenotype to ~omics data for bacteria: application to gene-trait matching for Lactobacillus plantarum strains. *BMC Genomics*. [Online] 13 (1), 170. Available from: doi:10.1093/bioinformatics/btn588.

de Been, M. *et al.* (2013) Recent recombination events in the core genome are associated with adaptive evolution in Enterococcus faecium. *Genome biology and evolution*. [Online] Available from: doi:10.1093/gbe/evt111.

Bentley, S.D. *et al.* (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS genetics*. [Online] 2 (3), e31. Available from: doi:10.1371/journal.pgen.0020031 [Accessed: 11 February 2015].

Bentley, S.D. & Parkhill, J. (2004) Comparative genomic structure of prokaryotes. *Annual review of genetics*. [Online] 38771–792. Available from: doi:10.1146/annurev.genet.38.072902.094318 [Accessed: 17 July 2015].

Berg, J.M., Tymoczko, J.L. & Stryer, L. (2002) *Hemoglobin Transports Oxygen Efficiently by Binding Oxygen Cooperatively*. [Online]. Available from:

http://www.ncbi.nlm.nih.gov/books/NBK22596/ [Accessed: 5 July 2015].

Bergin, D. *et al.* (2005) Superoxide Production in Galleria mellonella Hemocytes: Identification of Proteins Homologous to the NADPH Oxidase Complex of Human Neutrophils. *Infection and immunity*. [Online] 73 (7), 4161–4170. Available from: doi:10.1128/IAI.73.7.4161-4170.2005.

Bertels, F. *et al.* (2014) Automated reconstruction of whole genome phylogenies from short sequence reads. *Molecular biology and evolution*. [Online] Available from: doi:10.1093/molbev/msu088.

Best, E.L. *et al.* (2010) The potential for airborne dispersal of Clostridium difficile from symptomatic patients. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 50 (11), 1450–1457. Available from: doi:10.1086/652648 [Accessed: 15 June 2015].

Bignardi, G.E. (1998) Risk factors for Clostridium difficile infection. *Journal of Hospital Infection*. [Online] 40 (1), 1–15. Available from: doi:10.1016/S0195-6701(98)90019-6 [Accessed: 25 May 2015].

Blahna, M.T. (2006) The role of horizontal gene transfer in the spread of trimethoprim-sulfamethoxazole resistance among uropathogenic Escherichia coli in Europe and Canada. *Journal of Antimicrobial Chemotherapy*. [Online] 57 (4), 666–672. Available from: doi:10.1093/jac/dkl020 [Accessed: 5 April 2015].

Blair, J.M.A. & Piddock, L.J. V (2009) Structure, function and inhibition of RND efflux pumps in Gram-negative bacteria: an update. *Current opinion in microbiology*. [Online] 12 (5), 512–519. Available from: doi:10.1016/j.mib.2009.07.003 [Accessed: 30 July 2015].

Blázquez, J., Oliver, A. & Gómez-Gómez, J.-M. (2002) Mutation and evolution of antibiotic resistance: antibiotics as promoters of antibiotic resistance? *Current drug targets*. [Online] 3 (4), 345–349. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12102604 [Accessed: 26 September 2015].

Boinett, C.J. & Cain, A.K. (2014) Recombination: genomic mix ‘n’ match. *Nature reviews. Microbiology*. [Online] 12 (12), 795. Available from: doi:10.1038/nrmicro3388 [Accessed: 8 September 2015].

Bosch, T. *et al.* (2010) PFGE diversity within the methicillin-resistant Staphylococcus aureus clonal lineage ST398. *BMC microbiology*. [Online] 10 (1), 40. Available from: doi:10.1186/1471-2180-10-40 [Accessed: 1 June 2015].

Boto, L. (2010) Horizontal gene transfer in evolution: facts and challenges. *Proceedings. Biological sciences / The Royal Society*. [Online] 277 (1683), 819–827. Available from: doi:10.1098/rspb.2009.1679 [Accessed: 16 June 2015].

Bottacini, F. *et al.* (2010) Comparative genomics of the genus Bifidobacterium. *Microbiology (Reading, England)*. [Online] 156 (Pt 11), 3243–3254. Available from: doi:10.1099/mic.0.039545-0.

Boucher, H.W. *et al.* (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 48 (1), 1–12. Available from: doi:10.1086/595011 [Accessed: 9 July 2014].

Bouzid, M. *et al.* (2014) Climate change and the emergence of vector-borne diseases in Europe: case study of dengue fever. *BMC public health*. [Online] 14 (1), 781. Available from: doi:10.1186/1471-2458-14-781 [Accessed: 12 June 2015].

Brady, M.T. *et al.* (1989) Diagnostic studies of nosocomial diarrhea in child en: Assessing their use end value. *American Journal of Infection Control*. [Online] 17 (2), 77–82. Available from: doi:10.1016/0196-6553(89)90021-7 [Accessed: 15 June 2015].

Braker, G., Fesefeldt, A. & Witzel, K.-P. (1998) Development of PCR Primer Systems for Amplification of Nitrite Reductase Genes (nirK and nirS) To Detect Denitrifying Bacteria in Environmental Samples. *Appl. Envir. Microbiol.* [Online] 64 (10), 3769–3775. Available from: http://aem.asm.org/content/64/10/3769.full [Accessed: 6 August 2015].

Bratzler, D.W. & Hunt, D.R. (2006) The surgical infection prevention and surgical care improvement projects: national initiatives to improve outcomes for patients having surgery. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 43 (3), 322–330. Available from: doi:10.1086/505220 [Accessed: 21 August 2015].

Bravo, A. *et al.* (2011) Bacillus thuringiensis: A story of a successful bioinsecticide. *Insect biochemistry and molecular biology*. [Online] 41 (7), 423–431. Available from: doi:10.1016/j.ibmb.2011.02.006 [Accessed: 17 March 2015].

Breiman, L. (2001) Random Forests. *Machine Learning*. [Online] 45 (1), 5–32. Available from: doi:10.1023/A:1010933404324 [Accessed: 28 April 2014].

Brody, M.S., Vijay, K. & Price, C.W. (2001) Catalytic function of an alpha/beta hydrolase is required for energy stress activation of the sigma(B) transcription factor in Bacillus subtilis. *Journal of bacteriology*. [Online] 183 (21), 6422–6428. Available from: doi:10.1128/JB.183.21.6422-6428.2001 [Accessed: 14 August 2015].

Brown, N.F. *et al.* (2006) Crossing the line: selection and evolution of virulence traits. *PLoS pathogens*. [Online] 2 (5), e42. Available from: doi:10.1371/journal.ppat.0020042 [Accessed: 10 August 2015].

Brown, S.P., Cornforth, D.M. & Mideo, N. (2012) Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in microbiology*. [Online] 20 (7), 336–342. Available from: doi:10.1016/j.tim.2012.04.005 [Accessed: 3 June 2015].

Brueggemann, A.B. *et al.* (2007) Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS pathogens*. [Online] 3 (11), e168. Available from: doi:10.1371/journal.ppat.0030168 [Accessed: 12 June 2015].

Bryant, D. & Moulton, V. (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution*. [Online] 21 (2), 255–265. Available from: doi:10.1093/molbev/msh018.

Buer, J. & Balling, R. (2003) Mice, microbes and models of infection. *Nature reviews. Genetics*. [Online] 4 (3), 195–205. Available from: doi:10.1038/nrg1019 [Accessed: 21 April 2015].

Bullman, S. *et al.* (2013) Genomic investigation into strain heterogeneity and pathogenic potential of the emerging gastrointestinal pathogen Campylobacter ureolyticus. *PloS one*. [Online] 8 (8), e71515. Available from: doi:10.1371/journal.pone.0071515 [Accessed: 6 September 2015].

Burgos-Portugal, J.A. *et al.* (2012) Pathogenic potential of Campylobacter ureolyticus. *Infection and immunity*. [Online] 80 (2), 883–890. Available from: doi:10.1128/IAI.06031-11 [Accessed: 6 May 2015].

Burrus, V. & Waldor, M.K. (2004) Shaping bacterial genomes with integrative and conjugative elements. *Research in microbiology*. [Online] 155 (5), 376–386. Available from: doi:10.1016/j.resmic.2004.01.012 [Accessed: 17 August 2015].

Bush, K. (2010) Bench-to-bedside review: The role of beta-lactamases in antibiotic-resistant Gram-negative infections. *Critical care (London, England)*. [Online] 14 (3), 224. Available from: doi:10.1186/cc8892 [Accessed: 1 July 2014].

Canchaya, C. *et al.* (2003) Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*. [Online] 6 (4), 417–424. Available from: doi:10.1016/S1369-5274(03)00086-9 [Accessed: 17 August 2015].

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. [Online] 25 (15), 1972–1973. Available from: doi:10.1093/bioinformatics/btp348.

Carbonnelle, E. *et al.* (2011) MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clinical biochemistry*. [Online] 44 (1), 104–109. Available from: doi:10.1016/j.clinbiochem.2010.06.017 [Accessed: 18 May 2015].

Casadevall, A. & Pirofski, L. (2001) Host-pathogen interactions: the attributes of virulence. *The Journal of infectious diseases*. [Online] 184 (3), 337–344. Available from: doi:10.1086/322044 [Accessed: 8 August 2015].

Casadevall, A. & Pirofski, L. -a. (2000) Host-Pathogen Interactions: Basic Concepts of Microbial Commensalism, Colonization, Infection, and Disease. *Infection and Immunity*. [Online] 68 (12), 6511–6518. Available from: doi:10.1128/IAI.68.12.6511-6518.2000 [Accessed: 18 February 2015].

Castillo-Ramírez, S. *et al.* (2012) Phylogeographic variation in recombination rates within a global clone of methicillin-resistant Staphylococcus aureus. *Genome biology*. [Online] 13 (12), R126. Available from: doi:10.1186/gb-2012-13-12-r126.

Castillo-Ramírez, S. *et al.* (2011) The Impact of Recombination on dN/dS within Recently Emerged Bacterial Clones. *PLOS Pathogens*. [Online] 7 (7), e1002129. Available from: doi:10.1371/journal.ppat.1002129.

Chávez, V. *et al.* (2007) Oxidative stress enzymes are required for DAF-16-mediated immunity due to generation of reactive oxygen species by Caenorhabditis elegans. *Genetics*. [Online] 176 (3), 1567–1577. Available from: doi:10.1534/genetics.107.072587 [Accessed: 27 April 2015].

Chen, L. (2004) VFDB: a reference database for bacterial virulence factors. *Nucleic acids research*. [Online] 33 (Database issue), D325–D328. Available from: doi:10.1093/nar/gki008.

Chen, Y.-T. *et al.* (2012) Whole-genome sequencing and identification of Morganella morganii KT pathogenicity-related genes. *BMC genomics*. [Online] 13 Suppl 7S4. Available from: doi:10.1186/1471-2164-13-S7-S4.

Chen, Z. *et al.* (2014) IMP-1 encoded by a novel Tn402-like class 1 integron in clinical Achromobacter xylosoxidans, China. *Scientific reports*. [Online] 47212. Available from: doi:10.1038/srep07212 [Accessed: 8 September 2015].

Chewapreecha, C. *et al.* (2014) Dense genomic sampling identifies highways of pneumococcal recombination. *Nature Genetics*. [Online] -. Available from: doi:10.1038/ng.2895.

Chin, C.-S. *et al.* (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods*. [Online] 10563–569. Available from: doi:10.1038/nmeth.2474.

Chomel, B.B. & Kasten, R.W. (2010) Bartonellosis, an increasingly recognized zoonosis. *Journal of applied microbiology*. [Online] 109 (3), 743–750. Available from: doi:10.1111/j.1365-

2672.2010.04679.x [Accessed: 15 September 2015].

Choo, S.W. *et al.* (2014) Genomic reconnaissance of clinical isolates of emerging human pathogen Mycobacterium abscessus reveals high evolutionary potential. *Scientific reports*. [Online] 44061. Available from: doi:10.1038/srep04061 [Accessed: 23 June 2015].

Ciesielski, C.A. *et al.* (1988) The Geographic Distribution of Lyme Disease in the United States. *Annals of the New York Academy of Sciences*. [Online] 539 (1 Lyme Disease), 283–288. Available from: doi:10.1111/j.1749-6632.1988.tb31862.x [Accessed: 15 September 2015].

Ciofu, O., Hansen, C.R. & Høiby, N. (2013) Respiratory bacterial infections in cystic fibrosis. *Current Opinion in Pulmonary Medicine*. [Online] 19 (3), 251–258. Available from: doi:10.1097/MCP.0b013e32835f1afc.

Clatworthy, A.E. *et al.* (2009) Pseudomonas aeruginosa infection of zebrafish involves both host and pathogen determinants. *Infection and immunity*. [Online] 77 (4), 1293–1303. Available from: doi:10.1128/IAI.01181-08 [Accessed: 18 July 2015].

Coburn, B. *et al.* (2015) Lung microbiota across age and disease stage in cystic fibrosis. *Scientific Reports*. [Online] 510241. Available from: doi:10.1038/srep10241 [Accessed: 7 September 2015].

Coenye, T. *et al.* (2003) Achromobacter insolitus sp. nov. and Achromobacter spanius sp. nov., from human clinical samples. *International journal of systematic and evolutionary microbiology*. [Online] 53 (Pt 6), 1819–1824. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=14657110&retmode=ref&cmd=prlinks.

Coenye, T. *et al.* (2005) Advenella incenata gen. nov., sp. nov., a novel member of the Alcaligenaceae, isolated from various clinical samples. *International journal of systematic and evolutionary microbiology*. [Online] 55 (Pt 1), 251–256. Available from: doi:10.1099/ijs.0.63267-0 [Accessed: 27 May 2015].

Comas, I. *et al.* (2012) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nature genetics*. [Online] 44 (1), 106–110. Available from: doi:10.1038/ng.1038 [Accessed: 23 July 2015].

Conlan, S. *et al.* (2012) Staphylococcus epidermidis pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome biology*. [Online] 13 (7), R64. Available from: doi:10.1128/AEM.68.8.3673-3682.2002.

Croucher, N.J. *et al.* (2013) Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nature Genetics*. [Online] 45 (6), 656–663. Available from: doi:10.1038/ng.2625.

Croucher, N.J. *et al.* (2014) Variable recombination dynamics during the emergence, transmission and 'disarming' of a multidrug-resistant pneumococcal clone. *BMC biology*. [Online] 12 (1), 49. Available from: doi:10.1186/1741-7007-12-49 [Accessed: 2 July 2015].

Croxatto, A., Prod'hom, G. & Greub, G. (2012) Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiology Reviews*. [Online] 36 (2), 380–407. Available from: doi:10.1111/j.1574-6976.2011.00298.x.

Cutting, G.R. (2014) Cystic fibrosis genetics: from molecular understanding to clinical application. *Nature Reviews Genetics*. [Online] 16 (1), 45–56. Available from: doi:10.1038/nrg3849 [Accessed: 18 November 2014].

D'Argenio, D.A. *et al.* (2001) Drosophila as a model host for Pseudomonas aeruginosa infection.

Journal of bacteriology. [Online] 183 (4), 1466–1471. Available from: doi:10.1128/JB.183.4.1466-1471.2001 [Accessed: 16 September 2015].

Van Daele, S. *et al.* (2005) Shared genotypes of Achromobacter xylosoxidans strains isolated from patients at a cystic fibrosis rehabilitation center. *Journal of clinical microbiology*. [Online] 43 (6), 2998–3002. Available from: doi:10.1128/JCM.43.6.2998-3002.2005.

Dancer, S.J. (2014) Controlling hospital-acquired infection: focus on the role of the environment and new technologies for decontamination. *Clinical microbiology reviews*. [Online] 27 (4), 665–690. Available from: doi:10.1128/CMR.00020-14 [Accessed: 28 July 2015].

Danilova, N. *et al.* (2004) T cells and the thymus in developing zebrafish. *Developmental and comparative immunology*. [Online] 28 (7-8), 755–767. Available from: doi:10.1016/j.dci.2003.12.003 [Accessed: 27 April 2015].

Danilova, N. & Steiner, L.A. (2002) B cells develop in the zebrafish pancreas. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 99 (21), 13711–13716. Available from: doi:10.1073/pnas.212515999 [Accessed: 27 April 2015].

Danzmann, L. *et al.* (2013) Health care workers causing large nosocomial outbreaks: a systematic review. *BMC infectious diseases*. [Online] 13 (1), 98. Available from: doi:10.1186/1471-2334-13-98 [Accessed: 29 May 2015].

Darch, S.E. *et al.* (2015) Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Scientific Reports*. [Online] 57649. Available from: doi:10.1038/srep07649 [Accessed: 12 January 2015].

Dave, B.P. *et al.* (2013) Enhanced biodegradation of total polycyclic aromatic hydrocarbons (TPAHs) by marine halotolerant Achromobacter xylosoxidans using Triton X-100 and β-cyclodextrin - A microcosm approach. *Marine pollution bulletin*. [Online] Available from: doi:10.1016/j.marpolbul.2013.12.027.

Dave, J. *et al.* (2015) Trends in antibiotic susceptibility of enteric fever isolates in East London. *Travel medicine and infectious disease*. [Online] 13 (3), 230–234. Available from: doi:10.1016/j.tmaid.2015.04.001 [Accessed: 8 June 2015].

Davies, J. & O'Connor, S. (1978) Enzymatic modification of aminoglycoside antibiotics: 3-N-acetyltransferase with broad specificity that determines resistance to the novel aminoglycoside apramycin. *Antimicrobial agents and chemotherapy*. [Online] 14 (1), 69–72. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=352406&tool=pmcentrez&rendertype=abstract [Accessed: 19 February 2015].

Davies, J.C. & Rubin, B.K. (2007) Emerging and unusual gram-negative infections in cystic fibrosis. *Seminars in respiratory and critical care medicine*. [Online] 28 (3), 312–321. Available from: doi:10.1055/s-2007-981652 [Accessed: 24 August 2014].

Degand, N. *et al.* (2008) Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nonfermenting gram-negative bacilli isolated from cystic fibrosis patients. *Journal of clinical microbiology*. [Online] 46 (10), 3361–3367. Available from: doi:10.1128/JCM.00569-08.

Derber, C. *et al.* (2011) Achromobacter species endocarditis: A case report and literature review. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie médicale / AMMI Canada*. [Online] 22 (3), e17–e20. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22942890&

retmode=ref&cmd=prlinks.

Didelot, X. *et al.* (2012) Impact of homologous and non-homologous recombination in the genomic evolution of Escherichia coli. *BMC Genomics*. [Online] 13 (1), 256. Available from: doi:10.1186/1471-2164-13-256 [Accessed: 7 April 2015].

Didelot, X. & Falush, D. (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics*. [Online] 175 (3), 1251–1266. Available from: doi:10.1534/genetics.106.063305 [Accessed: 1 October 2014].

Didelot, X. & Wilson, D.J. (2015) ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS computational biology*. [Online] 11 (2), e1004041. Available from: doi:10.1371/journal.pcbi.1004041 [Accessed: 13 February 2015].

van Dijk, E.L. *et al.* (2014) Ten years of next-generation sequencing technology. *Trends in Genetics*. [Online] 30 (9), 418–426. Available from: doi:10.1016/j.tig.2014.07.001 [Accessed: 7 August 2014].

Dionne, M.S. & Schneider, D.S. (2008) Models of infectious diseases in the fruit fly Drosophila melanogaster. *Disease models & mechanisms*. [Online] 1 (1), 43–49. Available from: doi:10.1242/dmm.000307 [Accessed: 24 September 2015].

Djordjevic, Z. *et al.* (2013) Nosocomial urinary tract infections caused by Pseudomonas aeruginosa and Acinetobacter species: sensitivity to antibiotics and risk factors. *American journal of infection control*. [Online] 41 (12), 1182–1187. Available from: doi:10.1016/j.ajic.2013.02.018 [Accessed: 15 June 2015].

Dohm, J.C. *et al.* (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic acids research*. [Online] 36 (16), e105. Available from: doi:10.1093/nar/gkn425 [Accessed: 9 September 2015].

Doi, Y. *et al.* (2008) Characterization of a naturally occurring class D beta-lactamase from Achromobacter xylosoxidans. *Antimicrobial agents and chemotherapy*. [Online] 52 (6), 1952–1956. Available from: doi:10.1128/AAC.01463-07.

Domingues, S. *et al.* (2012) Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS pathogens*. [Online] 8 (8), e1002837. Available from: doi:10.1371/journal.ppat.1002837 [Accessed: 8 September 2015].

Donati, C. *et al.* (2010) Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome biology*. [Online] 11 (10), R107. Available from: doi:10.1186/gb-2010-11-10-r107.

Dong, J. *et al.* (2008) Emerging pathogens: challenges and successes of molecular diagnostics. *The Journal of molecular diagnostics : JMD*. [Online] 10 (3), 185–197. Available from: doi:10.2353/jmoldx.2008.070063 [Accessed: 27 August 2015].

Dowding, J.E. (1979) A novel aminoglycoside-modifying enzyme from a clinical isolate of Acinetobacter. *Journal of general microbiology*. [Online] 110 (1), 239–241. Available from: doi:10.1099/00221287-110-1-239 [Accessed: 8 September 2015].

Drlica, K. (1999) Mechanism of fluoroquinolone action. *Current Opinion in Microbiology*. [Online] 2 (5), 504–508. Available from: doi:10.1016/S1369-5274(99)00008-9 [Accessed: 18 February 2015].

Dubberke, E.R. & Olsen, M.A. (2012) Burden of Clostridium difficile on the healthcare system. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*.

[Online] 55 Suppl 2 (suppl_2), S88–S92. Available from: doi:10.1093/cid/cis335 [Accessed: 14 June 2015].

Dunne, W.M. & Maisch, S. (1995) Epidemiological investigation of infections due to Alcaligenes species in children and patients with cystic fibrosis: use of repetitive-element-sequence polymerase chain reaction. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 20 (4), 836–841. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=7795082&retmode=ref&cmd=prlinks.

Dunne, W.M., Westblade, L.F. & Ford, B. (2012) Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 31 (8), 1719–1726. Available from: doi:10.1007/s10096-012-1641-7.

Dutilh, B.E. *et al.* (2014) Comparative genomics of 274 Vibrio cholerae genomes reveals mobile functions structuring three niche dimensions. *BMC Genomics*. [Online] 15 (1), 654. Available from: doi:10.1186/1471-2164-15-654 [Accessed: 6 August 2014].

Eberl, M. *et al.* (2003) Microbial isoprenoid biosynthesis and human γδ T cell activation. *FEBS Letters*. [Online] 544 (1-3), 4–10. Available from: doi:10.1016/S0014-5793(03)00483-6 [Accessed: 17 February 2015].

Edgar, R.C. (2004) BMC Bioinformatics | Full text | MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. [Online] 5 (1), 113. Available from: doi:10.1186/1471-2105-5-113.

Egghe, L. (2007) Untangling Herdan's law and Heaps' law: Mathematical and informetric arguments. *Journal of the American Society for Information Science and Technology*. [Online] 58 (5), 702–709. Available from: doi:10.1002/asi.20524 [Accessed: 9 April 2015].

Eid, J. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)*. [Online] 323 (5910), 133–138. Available from: doi:10.1126/science.1162986 [Accessed: 13 July 2014].

Eliopoulos, G.M. & Huovinen, P. (2001) Resistance to trimethoprim-sulfamethoxazole. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 32 (11), 1608–1614. Available from: doi:10.1086/320532 [Accessed: 8 September 2015].

Elliot, S.L., Blanford, S. & Thomas, M.B. (2002) Host-pathogen interactions in a varying environment: temperature, behavioural fever and fitness. *Proceedings. Biological sciences / The Royal Society*. [Online] 269 (1500), 1599–1607. Available from: doi:10.1098/rspb.2002.2067 [Accessed: 25 August 2015].

Emerson, D. *et al.* (2008) Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics. *BioScience*. [Online] 58 (10), 925. Available from: doi:10.1641/B581006 [Accessed: 26 January 2015].

Emes, R.D. (2003) Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Human Molecular Genetics*. [Online] 12 (7), 701–709. Available from: doi:10.1093/hmg/ddg078 [Accessed: 19 March 2015].

Endtz, H.P. *et al.* (1991) Quinolone resistance in campylobacter isolated from man and poultry following the introduction of fluoroquinolones in veterinary medicine. *Journal of Antimicrobial Chemotherapy*. [Online] 27 (2), 199–208. Available from: doi:10.1093/jac/27.2.199 [Accessed: 2

April 2015].

Engel, P. *et al.* (2011) Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen Bartonella. *PLoS genetics*. [Online] 7 (2), e1001296. Available from: doi:10.1371/journal.pgen.1001296 [Accessed: 4 August 2015].

Engering, A., Hogerwerf, L. & Slingenbergh, J. (2013) Pathogen–host–environment interplay and disease emergence. *Emerging Microbes & Infections*. [Online] 2 (2), e5. Available from: doi:10.1038/emi.2013.5 [Accessed: 5 June 2015].

Enright, M.C. *et al.* (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. *Journal of clinical microbiology*. [Online] 38 (3), 1008–1015. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=86325&tool=pmcentrez&rendertype=abstract [Accessed: 6 September 2015].

Eppinger, M. *et al.* (2010) Genome sequence of the deep-rooted Yersinia pestis strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *Journal of bacteriology*. [Online] 192 (6), 1685–1699. Available from: doi:10.1128/JB.01518-09 [Accessed: 3 July 2015].

Erridge, C., Bennett-Guerrero, E. & Poxton, I.R. (2002) Structure and function of lipopolysaccharides. *Microbes and Infection*. [Online] 4 (8), 837–851. Available from: doi:10.1016/S1286-4579(02)01604-0 [Accessed: 3 February 2015].

Ewbank, J.J. & Zugasti, O. (2011) C. elegans: model host and tool for antimicrobial drug discovery. *Disease models & mechanisms*. [Online] 4 (3), 300–304. Available from: doi:10.1242/dmm.006684 [Accessed: 24 September 2015].

Falcone, M. *et al.* (2011) Healthcare-associated pneumonia: diagnostic criteria and distinction from community-acquired pneumonia. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*. [Online] 15 (8), e545–e550. Available from: doi:10.1016/j.ijid.2011.04.005 [Accessed: 13 June 2015].

Federici, L. *et al.* (2005) The Crystal Structure of the Outer Membrane Protein VceC from the Bacterial Pathogen Vibrio cholerae at 1.8 A Resolution. *Journal of Biological Chemistry*. [Online] 280 (15), 15307–15314. Available from: doi:10.1074/jbc.M500401200 [Accessed: 28 June 2015].

Feil, E.J. *et al.* (2004) eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *Journal of Bacteriology*. [Online] 186 (5), 1518–1530. Available from: doi:10.1128/JB.186.5.1518-1530.2004 [Accessed: 6 July 2015].

Feil, E.J. *et al.* (2003) How Clonal Is Staphylococcus aureus? *Journal of Bacteriology*. [Online] 185 (11), 3307–3316. Available from: doi:10.1128/JB.185.11.3307-3316.2003 [Accessed: 1 June 2015].

Feil, E.J. *et al.* (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 98 (1), 182–187. Available from: doi:10.1073/pnas.98.1.182 [Accessed: 7 April 2015].

Feng, Y. *et al.* (2014) A Francisella virulence factor catalyses an essential reaction of biotin synthesis. *Molecular microbiology*. [Online] 91 (2), 300–314. Available from: doi:10.1111/mmi.12460 [Accessed: 14 August 2015].

Fenselau, C. & Demirev, P.A. (1975) Characterization of intact microorganisms by MALDI mass spectrometry. *Mass spectrometry reviews*. [Online] 20 (4), 157–171. Available from: doi:10.1002/mas.10004 [Accessed: 28 May 2015].

Fernández-Olmos, A. *et al.* (2012) MALDI-TOF MS improves routine identification of non-fermenting Gram negative isolates from cystic fibrosis patients. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] 11 (1), 59–62. Available from: doi:10.1016/j.jcf.2011.09.001.

Ferroni, A. *et al.* (2002) Use of 16S rRNA gene sequencing for identification of nonfermenting gram-negative bacilli recovered from patients attending a single cystic fibrosis center. *Journal of clinical microbiology*. [Online] 40 (10), 3793–3797. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12354883&retmode=ref&cmd=prlinks.

Fischer, F., Künne, S. & Fetzner, S. (1999) Bacterial 2,4-dioxygenases: new members of the alpha/beta hydrolase-fold superfamily of enzymes functionally related to serine hydrolases. *Journal of bacteriology*. [Online] 181 (18), 5725–5733. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=94093&tool=pmcentrez&rendertype=abstract [Accessed: 14 August 2015].

Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (New York, N.Y.)*. [Online] 269 (5223), 496–512. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7542800 [Accessed: 18 December 2014].

Fleming, A. (1929) On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British journal of experimental pathology*. [Online]. 10 (3) p.226. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2048009/ [Accessed: 29 September 2015].

Fothergill, J.L. *et al.* (2010) Impact of Pseudomonas aeruginosa genomic instability on the application of typing methods for chronic cystic fibrosis infections. *Journal of clinical microbiology*. [Online] 48 (6), 2053–2059. Available from: doi:10.1128/JCM.00019-10 [Accessed: 1 June 2015].

Francisco, A.P. *et al.* (2009) Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC bioinformatics*. [Online] 10 (1), 152. Available from: doi:10.1186/1471-2105-10-152 [Accessed: 23 July 2015].

Frankel, G. (2001) Intimin and the host cell — is it bound to end in Tir(s)? *Trends in Microbiology*. [Online] 9 (5), 214–218. Available from: doi:10.1016/S0966-842X(01)02016-9 [Accessed: 3 August 2015].

Fricke, W.F. & Rasko, D.A. (2014) Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nature reviews. Genetics*. [Online] 15 (1), 49–55. Available from: doi:10.1038/nrg3624 [Accessed: 14 July 2014].

Friedman, S.M., Lu, T. & Drlica, K. (2001) Mutation in the DNA gyrase A Gene of Escherichia coli that expands the quinolone resistance-determining region. *Antimicrobial agents and chemotherapy*. [Online] 45 (8), 2378–2380. Available from: doi:10.1128/AAC.45.8.2378-2380.2001 [Accessed: 1 May 2015].

Frost, L.S. *et al.* (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*. [Online] 3 (9), 722–732. Available from: doi:10.1038/nrmicro1235 [Accessed: 9 July 2014].

Fujisaki, S. *et al.* (1996) Cloning of a gene from Escherichia coli that confers resistance to fosmidomycin as a consequence of amplification. *Gene*. [Online] 175 (1-2), 83–87. Available from: doi:10.1016/0378-1119(96)00128-X [Accessed: 17 February 2015].

Gaillot, O. *et al.* (2011) Cost-effectiveness of switch to matrix-assisted laser desorption ionization-time of flight mass spectrometry for routine bacterial identification. *Journal of clinical microbiology*. [Online] 49 (12), 4412. Available from: doi:10.1128/JCM.05429-11 [Accessed: 1 June 2015].

Gens, K.D., Elshaboury, R.H. & Holt, J.S. (2013) Fecal microbiota transplantation and emerging treatments for Clostridium difficile infection. *Journal of pharmacy practice*. [Online] 26 (5), 498–505. Available from: doi:10.1177/0897190013499527 [Accessed: 27 September 2015].

Gerhardt, P. (1994) *Methods for General and Molecular Bacteriology*. [Online]. Washington D.C., American Society for Microbiology. Available from: http://books.google.co.uk/books/about/Methods_for_General_and_Molecular_Bacter.html?id=fGYXAQAAIAAJ&pgis=1 [Accessed: 16 December 2014].

Giacoia, G.P. (1990) Achromobacter xylosoxidans: a drug-resistant pathogen in newborns and immunocompromised patients. *Southern medical journal*. [Online] 83 (11), 1312–1314. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=2237562&retmode=ref&cmd=prlinks.

Gibson, R.L., Burns, J.L. & Ramsey, B.W. (2003) Pathophysiology and management of pulmonary infections in cystic fibrosis. *American journal of respiratory and critical care medicine*. [Online] 168 (8), 918–951. Available from: doi:10.1164/rccm.200304-505SO.

Gilbert, R.J., Stringer, M.F. & Peace, T.C. (1974) The survival and growth of Bacillus cereus in boiled and fried rice in relation to outbreaks of food poisoning. *The Journal of hygiene*. [Online] 73 (3), 433–444. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2130471&tool=pmcentrez&rendertype=abstract [Accessed: 7 April 2015].

Gillings, M. *et al.* (2008) The Evolution of Class 1 Integrons and the Rise of Antibiotic Resistance. *Journal of bacteriology*. [Online] 190 (14), 5095–5100. Available from: doi:10.1128/JB.00152-08.

Gillings, M.R. *et al.* (2009) Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? *The ISME journal*. [Online] 3 (2), 209–215. Available from: doi:10.1038/ismej.2008.98 [Accessed: 28 July 2015].

Gladman, S. & Seemann, T. (2008) *VelvetOptimiser*. [Online]. Available from: http://bioinformatics.net.au/software.velvetoptimiser.shtml.

Goering, R. V (2010) Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. [Online] 10 (7), 866–875. Available from: doi:10.1016/j.meegid.2010.07.023 [Accessed: 1 June 2015].

Gold, H.S. & Moellering, R.C. (1996) Antimicrobial-drug resistance. *The New England journal of medicine*. [Online] 335 (19), 1445–1453. Available from: doi:10.1056/NEJM199611073351907 [Accessed: 16 September 2015].

Gómez-Cerezo, J. *et al.* (2003) Achromobacter xylosoxidans bacteremia: a 10-year analysis of 54 cases. *European journal of clinical microbiology & infectious diseases : official publication*

of the European Society of Clinical Microbiology. [Online] 22 (6), 360–363. Available from: doi:10.1007/s10096-003-0925-3.

Gomila, M. *et al.* (2014) Genotypic and phenotypic applications for the differentiation and species-level identification of achromobacter for clinical diagnoses. *PloS one*. [Online] 9 (12), e114356. Available from: doi:10.1371/journal.pone.0114356 [Accessed: 5 January 2015].

González, L.J. & Vila, A.J. (2012) Carbapenem resistance in Elizabethkingia meningoseptica is mediated by metallo-β-lactamase BlaB. *Antimicrobial agents and chemotherapy*. [Online] 56 (4), 1686–1692. Available from: doi:10.1128/AAC.05835-11 [Accessed: 18 September 2015].

Gordon, N.C. *et al.* (2014) Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. *Journal of clinical microbiology*. [Online] 52 (4), 1182–1191. Available from: doi:10.1128/JCM.03117-13 [Accessed: 26 May 2015].

Grad, Y.H. & Lipsitch, M. (2014) Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome biology*. [Online] 15 (11), 538. Available from: doi:10.1186/s13059-014-0538-4 [Accessed: 20 June 2015].

Gribun, A. *et al.* (2003) Molecular and structural characterization of the HMP-AB gene encoding a pore-forming protein from a clinical isolate of Acinetobacter baumannii. *Current microbiology*. [Online] 47 (5), 434–443. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14669924 [Accessed: 8 September 2015].

Gueimonde, M. *et al.* (2010) Genetic basis of tetracycline resistance in Bifidobacterium animalis subsp. lactis. *Applied and environmental microbiology*. [Online] 76 (10), 3364–3369. Available from: doi:10.1128/AEM.03096-09 [Accessed: 22 June 2015].

Guindon, S. *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*. [Online] 59 (3), 307–321. Available from: doi:10.1093/sysbio/syq010 [Accessed: 9 July 2014].

Haldorsen, B.C. *et al.* (2014) Increased prevalence of aminoglycoside resistance in clinical isolates of Escherichia coli and Klebsiella spp. in Norway is associated with the acquisition of AAC(3)-II and AAC(6′)-Ib. *Diagnostic Microbiology and Infectious Disease*. [Online] 78 (1), 66–69. Available from: doi:10.1016/j.diagmicrobio.2013.10.001 [Accessed: 5 April 2015].

Haller, S. *et al.* (2015) What caused the outbreak of ESBL-producing Klebsiella pneumoniae in a neonatal intensive care unit, Germany 2009 to 2012? Reconstructing transmission with epidemiological analysis and whole-genome sequencing. *BMJ open*. [Online] 5 (5), e007397. Available from: doi:10.1136/bmjopen-2014-007397 [Accessed: 20 May 2015].

Hallett, P. & Maxwell, A. (1991) Novel quinolone resistance mutations of the Escherichia coli DNA gyrase A protein: enzymatic analysis of the mutant proteins. *Antimicrobial agents and chemotherapy*. [Online] 35 (2), 335–340. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=245001&tool=pmcentrez&rendertype=abstract [Accessed: 20 February 2015].

Hanage, W.P. *et al.* (2009) Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science (New York, N.Y.)*. [Online] 324 (5933), 1454–1457. Available from: doi:10.1126/science.1171908 [Accessed: 20 August 2015].

Hansen, C.R. *et al.* (2013) Achromobacter species in cystic fibrosis: Cross-infection caused by indirect patient-to-patient contact. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] 2 (6), 609–615. Available from: doi:10.1016/j.jcf.2013.05.004.

Hansen, C.R. *et al.* (2010) Inflammation in Achromobacter xylosoxidans infected cystic fibrosis

patients. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] 9 (1), 51–58. Available from: doi:10.1016/j.jcf.2009.10.005.

Hansson, K. *et al.* (2002) IntI2 integron integrase in Tn7. *Journal of bacteriology*. [Online] 184 (6), 1712–1721. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=134885&tool=pmcentrez&rendertype=abstract [Accessed: 28 July 2015].

Hardison, R.C. (2012) Evolution of hemoglobin and its genes. *Cold Spring Harbor perspectives in medicine*. [Online] 2 (12), a011627. Available from: doi:10.1101/cshperspect.a011627 [Accessed: 5 July 2015].

Harris, S.R. *et al.* (2010) Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science*. [Online] 327 (5964), 469–474. Available from: doi:10.1126/science.1182395.

Harris, S.R. *et al.* (2013) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *The Lancet infectious diseases*. [Online] 13 (2), 130–136. Available from: doi:10.1016/S1473-3099(12)70268-2.

Hartman, B.J. & Tomasz, A. (1984) Low-affinity penicillin-binding protein associated with beta-lactam resistance in Staphylococcus aureus. *J. Bacteriol.* [Online] 158 (2), 513–516. Available from: http://jb.asm.org/content/158/2/513.abstract?ijkey=d26bc78345620445ba3e01bcf61bc6e0b871721c&keytype2=tf_ipsecsha [Accessed: 2 April 2015].

Hawker, J.I. *et al.* (2014) Trends in antibiotic prescribing in primary care for clinical syndromes subject to national recommendations to reduce antibiotic resistance, UK 1995-2011: analysis of a large database of primary care consultations. *The Journal of antimicrobial chemotherapy*. [Online] 69 (12), 3423–3430. Available from: doi:10.1093/jac/dku291 [Accessed: 27 July 2015].

Heddle, J. & Maxwell, A. (2002) Quinolone-Binding Pocket of DNA Gyrase: Role of GyrB. *Antimicrobial Agents and Chemotherapy*. [Online] 46 (6), 1805–1815. Available from: doi:10.1128/AAC.46.6.1805-1815.2002 [Accessed: 20 February 2015].

Hedges, S.B. (2002) The origin and evolution of model organisms. *Nature reviews. Genetics*. [Online] 3 (11), 838–849. Available from: doi:10.1038/nrg929 [Accessed: 20 April 2015].

Hochhut, B. *et al.* (2001) Molecular analysis of antibiotic resistance gene clusters in vibrio cholerae O139 and O1 SXT constins. *Antimicrobial agents and chemotherapy*. [Online] 45 (11), 2991–3000. Available from: doi:10.1128/AAC.45.11.2991-3000.2001 [Accessed: 28 July 2015].

Hogg, J.S. *et al.* (2007) Characterization and modeling of the Haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome biology*. [Online] 8 (6), R103. Available from: doi:10.1186/gb-2007-8-6-r103 [Accessed: 2 July 2015].

Holmes, B., Snell, J.J. & Lapage, S.P. (1977) Strains of Achromobacter xylosoxidans from clinical material. *Journal of clinical pathology*. [Online] 30 (7), 595–601. Available from: doi:10.1136/jcp.30.7.595.

Hooton, T.M. *et al.* (2010) Diagnosis, Prevention, and Treatment of Catheter-Associated Urinary Tract Infection in Adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America. *Clinical Infectious Diseases*. [Online] 50 (5), 625–663. Available from: doi:10.1086/650482 [Accessed: 11 July 2014].

Hornsey, M. *et al.* (2013) In vivo efficacy of telavancin/colistin combination therapy in a Galleria mellonella model of Acinetobacter baumannii infection. *International journal of antimicrobial*

agents. [Online] 41 (3), 285–287. Available from: doi:10.1016/j.ijantimicag.2012.11.013.

Howard, D.H. (2004) Resistance-induced antibiotic substitution. *Health economics*. [Online] 13 (6), 585–595. Available from: doi:10.1002/hec.856 [Accessed: 18 June 2015].

Howell, K.J. *et al.* (2014) The use of genome wide association methods to investigate pathogenicity, population structure and serovar in Haemophilus parasuis. *BMC genomics*. [Online] 15 (1), 1179. Available from: doi:10.1186/1471-2164-15-1179 [Accessed: 31 July 2015].

Hsueh, P.-R., Chen, W.-H. & Luh, K.-T. (2005) Relationships between antimicrobial use and antimicrobial resistance in Gram-negative bacteria causing nosocomial infections from 1991-2003 at a university hospital in Taiwan. *International journal of antimicrobial agents*. [Online] 26 (6), 463–472. Available from: doi:10.1016/j.ijantimicag.2005.08.016 [Accessed: 19 June 2015].

Hu, Y. *et al.* (2015) Genomic insights into intrinsic and acquired drug resistance mechanisms in Achromobacter xylosoxidans. *Antimicrobial agents and chemotherapy*. [Online] 59 (2), 1152–1161. Available from: doi:10.1128/AAC.04260-14 [Accessed: 7 September 2015].

Hunter, P. (2008) The paradox of model organisms. The use of model organisms in research will continue despite their shortcomings. *EMBO reports*. [Online] 9 (8), 717–720. Available from: doi:10.1038/embor.2008.142 [Accessed: 10 August 2015].

Huovinen, P. *et al.* (1995) Trimethoprim and sulfonamide resistance. *Antimicrobial agents and chemotherapy*. [Online] 39 (2), 279–289. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=162528&tool=pmcentrez&rendertype=abstract [Accessed: 27 February 2015].

Hyams, C. *et al.* (2010) The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infection and immunity*. [Online] 78 (2), 704–715. Available from: doi:10.1128/IAI.00881-09 [Accessed: 22 July 2015].

Hyatt, D. *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*. [Online] 11 (1), 119. Available from: doi:10.1186/1471-2105-11-119 [Accessed: 24 November 2014].

Hyodo, S., Katahira, S. & Shigeta, S. (1982) Experimental infection of immunocompromised mice with Achromobacter xylosoxidans. *The Tohoku journal of experimental medicine*. [Online] 136 (3), 251–261. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=6803396&retmode=ref&cmd=prlinks.

Igra-Siegman, Y., Chmel, H. & Cobbs, C. (1980) Clinical and laboratory characteristics of Achromobacter xylosoxidans infection. *Journal of clinical microbiology*. [Online] 11 (2), 141–145. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=7358838&retmode=ref&cmd=prlinks.

Ivanova, N. *et al.* (2003) Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis. *Nature*. [Online] 423 (6935), 87–91. Available from: doi:10.1038/nature01582 [Accessed: 7 April 2015].

Jacobsen, A. *et al.* (2011) The Salmonella enterica pan-genome. *Microbial ecology*. [Online] 62 (3), 487–504. Available from: doi:10.1007/s00248-011-9880-1 [Accessed: 31 July 2015].

Jain, S. & Ohman, D.E. (2005) Role of an alginate lyase for alginate transport in mucoid Pseudomonas aeruginosa. *Infection and immunity*. [Online] 73 (10), 6429–6436. Available from:

doi:10.1128/IAI.73.10.6429-6436.2005 [Accessed: 30 April 2015].

Jakobsen, T.H. *et al.* (2013) Complete Genome Sequence of the Cystic Fibrosis Pathogen Achromobacter xylosoxidans NH44784-1996 Complies with Important Pathogenic Phenotypes. *PLOS ONE*. [Online] 8 (7), e68484. Available from: doi:10.1371/journal.pone.0068484.

Jalal, S. *et al.* (2000) Molecular mechanisms of fluoroquinolone resistance in Pseudomonas aeruginosa isolates from cystic fibrosis patients. *Antimicrobial agents and chemotherapy*. [Online] 44 (3), 710–712. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=89751&tool=pmcentrez&rendertype=abstract [Accessed: 20 February 2015].

Janda, J.M. & Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology*. [Online] 45 (9), 2761–2764. Available from: doi:10.1128/JCM.01228-07 [Accessed: 9 July 2014].

Jander, G., Rahme, L.G. & Ausubel, F.M. (2000) Positive correlation between virulence of Pseudomonas aeruginosa mutants in mice and insects. *Journal of bacteriology*. [Online] 182 (13), 3843–3845. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10851003&retmode=ref&cmd=prlinks.

Jean, S.S. *et al.* (2014) Elizabethkingia meningoseptica: an important emerging pathogen causing healthcare-associated infections. *The Journal of hospital infection*. [Online] 86 (4), 244–249. Available from: doi:10.1016/j.jhin.2014.01.009 [Accessed: 26 April 2015].

Jensen, L.J. *et al.* (2008) eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic acids research*. [Online] 36 (Database issue), D250–D254. Available from: doi:10.1093/nar/gkm796 [Accessed: 28 September 2015].

Jones, R.N. (2010) Microbial etiologies of hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 51 Suppl 1S81–S87. Available from: doi:10.1086/653053 [Accessed: 11 June 2015].

Joshi, N.A. & Fass, J.N. (2011) *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)*. [Online]. Available from: https://github.com/najoshi/sickle.

Juretschko, S., Beavers-May, T.K. & Stovall, S.H. (2010) Nosocomial infection with Asaia lannensis in two paediatric patients with idiopathic dilated cardiomyopathy. *Journal of medical microbiology*. [Online] 59 (Pt 7), 848–852. Available from: doi:10.1099/jmm.0.019067-0 [Accessed: 19 June 2015].

Kadioglu, A. *et al.* (2008) The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nature reviews. Microbiology*. [Online] 6 (4), 288–301. Available from: doi:10.1038/nrmicro1871 [Accessed: 29 December 2014].

Kalanuria, A.A., Zai, W. & Mirski, M. (2014) Ventilator-associated pneumonia in the ICU. *Critical care (London, England)*. [Online] 18 (2), 208. Available from: doi:10.1186/cc13775 [Accessed: 18 April 2015].

Kamada, M. *et al.* (2014) Whole genome complete resequencing of Bacillus subtilis natto by combining long reads with high-quality short reads. *PloS one*. [Online] 9 (10), e109999. Available from: doi:10.1371/journal.pone.0109999 [Accessed: 27 August 2015].

Kanellopoulou, M. *et al.* (2004) Persistent colonization of nine cystic fibrosis patients with an Achromobacter (Alcaligenes) xylosoxidans clone. *European journal of clinical microbiology*

& infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 23 (4), 336–339. Available from: doi:10.1007/s10096-004-1105-9.

Kansakar, P. *et al.* (2011) Local dissemination of multidrug-resistant Acinetobacter baumannii clones in a Thai hospital. *Microbial drug resistance (Larchmont, N.Y.)*. [Online] 17 (1), 109–119. Available from: doi:10.1089/mdr.2010.0062.

Kargar, M. *et al.* (2014) High Prevalence of Class 1 to 3 Integrons Among Multidrug-Resistant Diarrheagenic Escherichia coli in Southwest of Iran. *Osong public health and research perspectives*. [Online] 5 (4), 193–198. Available from: doi:10.1016/j.phrp.2014.06.003 [Accessed: 28 July 2015].

Katzung, B.G., Masters, S.B. & Trevor, A.J. (2009) *Basic and Clinical Pharmacology*. [Online]. Available from: http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0071451536.

Kaur, M. *et al.* (2009) Epidemiological typing of clinical isolates of Achromobacter xylosoxidans: comparison of phenotypic and genotypic methods. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 28 (9), 1023–1032. Available from: doi:10.1007/s10096-009-0740-6.

Kiddee, A. *et al.* (2013) Nosocomial spread of class 1 integron-carrying extensively drug-resistant Pseudomonas aeruginosa isolates in a Thai hospital. *International journal of antimicrobial agents*. [Online] 42 (4), 301–306. Available from: doi:10.1016/j.ijantimicag.2013.05.009.

Kienesberger, S. *et al.* (2014) Comparative genome analysis of Campylobacter fetus subspecies revealed horizontally acquired genetic elements important for virulence and niche specificity. *PloS one*. [Online] 9 (1), e85491. Available from: doi:10.1371/journal.pone.0085491.

Kim, M.J. *et al.* (2013) Dissemination of metallo-β-lactamase-producing Pseudomonas aeruginosa of sequence type 235 in Asian countries. *The Journal of antimicrobial chemotherapy*. [Online] 68 (12), 2820–2824. Available from: doi:10.1093/jac/dkt269.

Kiser, K.B. *et al.* (1999) Staphylococcus aureus cap5P encodes a UDP-N-acetylglucosamine 2-epimerase with functional redundancy. *Journal of bacteriology*. [Online] 181 (16), 4818–4824. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=93967&tool=pmcentrez&rendertype=abstract [Accessed: 2 August 2015].

Kobayashi, N., Nishino, K. & Yamaguchi, A. (2001) Novel macrolide-specific ABC-type efflux transporter in Escherichia coli. *Journal of bacteriology*. [Online] 183 (19), 5639–5644. Available from: doi:10.1128/JB.183.19.5639-5644.2001 [Accessed: 5 May 2015].

Kolář, M., Urbánek, K. & Látal, T. (2001) Antibiotic selective pressure and development of bacterial resistance. *International Journal of Antimicrobial Agents*. [Online] 17 (5), 357–363. Available from: doi:10.1016/S0924-8579(01)00317-X [Accessed: 19 June 2015].

Kong, K.-F., Schneper, L. & Mathee, K. (2010) Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. [Online] 118 (1), 1–36. Available from: doi:10.1111/j.1600-0463.2009.02563.x [Accessed: 19 February 2015].

Koonin, E. V (2005) Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics*. [Online] 39309–338. Available from: doi:10.1146/annurev.genet.39.073003.114725 [Accessed: 10 July 2014].

Kos, V.N. *et al.* (2015) The Resistome of Pseudomonas aeruginosa in Relationship to Phenotypic Susceptibility. *Antimicrobial agents and chemotherapy*. [Online] 59 (1), 427–436. Available from: doi:10.1128/AAC.03954-14 [Accessed: 5 January 2015].

Köser, C.U. *et al.* (2012) Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLOS Pathogens*. [Online] 8 (8), e1002824. Available from: doi:10.1371/journal.ppat.1002824 [Accessed: 16 September 2014].

Köser, C.U., Ellington, M.J. & Peacock, S.J. (2014) Whole-genome sequencing to control antimicrobial resistance. *Trends in genetics : TIG*. [Online] 30 (9), 401–407. Available from: doi:10.1016/j.tig.2014.07.003 [Accessed: 25 March 2015].

Krishnamoorthy, G., Tikhonova, E.B. & Zgurskaya, H.I. (2008) Fitting periplasmic membrane fusion proteins to inner membrane transporters: mutations that enable Escherichia coli AcrA to function with Pseudomonas aeruginosa MexB. *Journal of bacteriology*. [Online] 190 (2), 691–698. Available from: doi:10.1128/JB.01276-07 [Accessed: 30 July 2015].

Krzewinski, J.W. *et al.* (2001) Use of random amplified polymorphic DNA PCR to examine epidemiology of Stenotrophomonas maltophilia and Achromobacter (Alcaligenes) xylosoxidans from patients with cystic fibrosis. *Journal of clinical microbiology*. [Online] 39 (10), 3597–3602. Available from: doi:10.1128/JCM.39.10.3597-3602.2001.

Kusuma Na Ayuthya, S. *et al.* (2003) *Utilization of restricted antibiotics in a university hospital in Thailand.* [Online] Available from: http://imsear.hellis.org/handle/123456789/30958.

Kwok, A.H. *et al.* (2014) Complete genome assembly and characterization of an outbreak strain of the causative agent of swine erysipelas--Erysipelothrix rhusiopathiae SY1027. *BMC microbiology*. 14176.

Laabei, M. *et al.* (2014) Predicting the virulence of MRSA from its genome sequence. *Genome research*. [Online] 24 (5), 839–849. Available from: doi:10.1101/gr.165415.113 [Accessed: 19 September 2015].

Laehnemann, D., Borkhardt, A. & McHardy, A.C. (2015) Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Briefings in bioinformatics*. [Online] Available from: doi:10.1093/bib/bbv029 [Accessed: 16 September 2015].

Lagesen, K. *et al.* (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*. [Online] 35 (9), 3100–3108. Available from: doi:10.1093/nar/gkm160 [Accessed: 25 November 2014].

Lambiase, A. *et al.* (2011) Achromobacter xylosoxidans respiratory tract infection in cystic fibrosis patients. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 30 (8), 973–980. Available from: doi:10.1007/s10096-011-1182-5.

Lau, S.K.P. *et al.* (2002) Identification by 16S ribosomal RNA gene sequencing of Arcobacter butzleri bacteraemia in a patient with acute gangrenous appendicitis. *Molecular pathology : MP*. [Online] 55 (3), 182–185. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1187171&tool=pmcentrez&rendertype=abstract [Accessed: 30 May 2015].

LeBlanc, J.G. *et al.* (2013) Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current opinion in biotechnology*. [Online] 24 (2), 160–168. Available from: doi:10.1016/j.copbio.2012.08.005 [Accessed: 10 May 2015].

Lebreton, F. *et al.* (2011) Galleria mellonella as a model for studying Enterococcus faecium host

persistence. *Journal of molecular microbiology and biotechnology*. [Online] 21 (3-4), 191–196. Available from: doi:10.1159/000332737.

Lee, W.-S. *et al.* (2014) Arteriovenous fistula-related bacteremia caused by Achromobacter xylosoxidans in a uremic patient. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi*. [Online] Available from: doi:10.1016/j.jmii.2014.03.011 [Accessed: 8 September 2015].

Leekitcharoenphon, P. *et al.* (2014) Evaluation of Whole Genome Sequencing for Outbreak Detection of Salmonella enterica. *PLoS ONE*. [Online] 9e87991. Available from: doi:10.1371/journal.pone.0087991.

Lery, L.M. *et al.* (2014) Comparative analysis of Klebsiella pneumoniae genomes identifies a phospholipase D family protein as a novel virulence factor. *BMC biology*. [Online] 12 (1), 41. Available from: doi:10.1186/1741-7007-12-41.

Levene, M.J. *et al.* (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. *Science (New York, N.Y.)*. [Online] 299 (5607), 682–686. Available from: doi:10.1126/science.1079700 [Accessed: 17 July 2015].

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. [Online] 25 (14), 1754–1760. Available from: doi:10.1093/bioinformatics/btp324.

Li, L., Stoeckert, C.J. & Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*. [Online] 13 (9), 2178–2189. Available from: doi:10.1101/gr.1224503.

Li, W., Raoult, D. & Fournier, P.-E. (2009) Bacterial strain typing in the genomic era. *FEMS microbiology reviews*. [Online] 33 (5), 892–916. Available from: doi:10.1111/j.1574-6976.2009.00182.x [Accessed: 11 March 2015].

Li, X. *et al.* (2013) Comparative genome characterization of Achromobacter members reveals potential genetic determinants facilitating the adaptation to a pathogenic lifestyle. *Applied microbiology and biotechnology*. [Online] 1–13. Available from: doi:10.1007/s00253-013-5018-3.

Lipsitch, M. & Moxon, E.R. (1997) Virulence and transmissibility of pathogens: what is the relationship? *Trends in microbiology*. [Online] 5 (1), 31–37. Available from: doi:10.1016/S0966-842X(97)81772-6 [Accessed: 9 August 2015].

Lipuma, J.J. (2010) The changing microbial epidemiology in cystic fibrosis. *Clinical microbiology reviews*. [Online] 23 (2), 299–323. Available from: doi:10.1128/CMR.00068-09 [Accessed: 31 July 2014].

Liu, F. *et al.* (2014) Comparative genomic analysis of Acinetobacter baumannii clinical isolates reveals extensive genomic variation and diverse antibiotic resistance determinants. *BMC Genomics*. [Online] 15 (1), 1163. Available from: doi:10.1186/1471-2164-15-1163 [Accessed: 7 September 2015].

Liu, L. *et al.* (2002) Ribosomal DNA-directed PCR for identification of Achromobacter (Alcaligenes) xylosoxidans recovered from sputum samples from cystic fibrosis patients. *Journal of clinical microbiology*. [Online] 40 (4), 1210–1213. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=11923333&retmode=ref&cmd=prlinks.

Loman, N.J. *et al.* (2012) High-throughput bacterial genome sequencing: an embarrassment of

choice, a world of opportunity. *Nature reviews. Microbiology*. [Online] 10 (9), 599–606. Available from: doi:10.1038/nrmicro2850 [Accessed: 16 July 2014].

Looney, W.J., Narita, M. & Mühlemann, K. (2009) Stenotrophomonas maltophilia: an emerging opportunist human pathogen. *The Lancet. Infectious diseases*. [Online] 9 (5), 312–323. Available from: doi:10.1016/S1473-3099(09)70083-0 [Accessed: 4 June 2015].

Losada, L. *et al.* (2010) Continuing evolution of Burkholderia mallei through genome reduction and large-scale rearrangements. *Genome biology and evolution*. [Online] 2102–116. Available from: doi:10.1093/gbe/evq003 [Accessed: 17 August 2015].

Lunetta, K.L. *et al.* (2004) Screening large-scale association study data: exploiting interactions using random forests. *BMC genetics*. [Online] 5 (1), 32. Available from: doi:10.1186/1471-2156-5-32 [Accessed: 18 July 2015].

Lux, T.M., Lee, R. & Love, J. (2014) GENOME-WIDE PHYLOGENETIC ANALYSIS OF THE PATHOGENIC POTENTIAL OF VIBRIO FURNISSII. *Aquatic Microbiology*. [Online] 5. Available from: doi:10.3389/fmicb.2014.00435 [Accessed: 11 August 2014].

Lynch, J.P., Clark, N.M. & Zhanel, G.G. (2013) Evolution of antimicrobial resistance among Enterobacteriaceae (focus on extended spectrum β-lactamases and carbapenemases). *Expert opinion on pharmacotherapy*. [Online] 14 (2), 199–210. Available from: doi:10.1517/14656566.2013.763030 [Accessed: 27 August 2015].

MacConkey, A.T. (1900) NOTE ON A NEW MEDIUM FOR THE GROWTH AND DIFFERENTIATION OF THE BACILLUS COLI COMMUNIS AND THE BACILLUS TYPHI ABDOMINALIS. *The Lancet*. [Online] 156 (4010), 20. Available from: doi:10.1016/S0140-6736(01)99513-3 [Accessed: 29 May 2015].

Magill, S.S. *et al.* (2014) Multistate point-prevalence survey of health care-associated infections. *The New England journal of medicine*. [Online] 370 (13), 1198–1208. Available from: doi:10.1056/NEJMoa1306801 [Accessed: 10 May 2015].

Magni, A. *et al.* (2010) Achromobacter xylosoxidans genomic characterization and correlation of randomly amplified polymorphic DNA profiles with relevant clinical features [corrected] of cystic fibrosis patients. *Journal of clinical microbiology*. [Online] 48 (4), 1035–1039. Available from: doi:10.1128/JCM.02060-09.

Mahenthiralingam, E. (2014) Emerging cystic fibrosis pathogens and the microbiome. *Paediatric Respiratory Reviews*. [Online] 15 (S1), 13–15. Available from: doi:10.1016/j.prrv.2014.04.006.

Mahenthiralingam, E. *et al.* (1996) Random amplified polymorphic DNA typing of Pseudomonas aeruginosa isolates recovered from patients with cystic fibrosis. *Journal of clinical microbiology*. [Online] 34 (5), 1129–1135. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=228968&tool=pmcentrez&rendertype=abstract [Accessed: 18 May 2015].

Maiden, M.C. (1998) Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 27 Suppl 1S12–S20. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9710667&retmode=ref&cmd=prlinks.

Maiden, M.C. *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 95 (6), 3140–3145. Available

from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=19708&tool=pmcentrez&rendertype=abstract [Accessed: 15 May 2015].

Manian, F.A. *et al.* (2003) Surgical site infections associated with methicillin-resistant Staphylococcus aureus: do postoperative factors play a role? *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 36 (7), 863–868. Available from: doi:10.1086/368195 [Accessed: 21 August 2015].

Mantovani, R.P. *et al.* (2010) Virulence properties of Achromobacter xylosoxidans: cytotoxic activity, interleukin production, biofilm formation and invasion. *Journal of Cystic Fibrosis*. [Online] 9S35. Available from: doi:10.1016/S1569-1993(10)60134-8 [Accessed: 2 August 2015].

Mantovani, R.P., Levy, C.E. & Yano, T. (2012) A heat-stable cytotoxic factor produced by Achromobacter xylosoxidans isolated from Brazilian patients with CF is associated with in vitro increased proinflammatory cytokines. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] Available from: doi:10.1016/j.jcf.2012.02.002.

Mardis, E.R. (2013) Next-generation sequencing platforms. *Annual review of analytical chemistry (Palo Alto, Calif.)*. [Online] 6287–303. Available from: doi:10.1146/annurev-anchem-062012-092628 [Accessed: 9 July 2014].

Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. [Online] 437 (7057), 376–380. Available from: doi:10.1038/nature03959 [Accessed: 10 July 2014].

Marra, A.R. *et al.* (2007) Hospital-acquired Clostridium difficile-associated disease in the intensive care unit setting: epidemiology, clinical course and outcome. *BMC infectious diseases*. [Online] 7 (1), 42. Available from: doi:10.1186/1471-2334-7-42 [Accessed: 6 September 2015].

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. [Online] 17 (1), 10. Available from: doi:10.14806/ej.17.1.200 [Accessed: 12 January 2015].

Marttinen, P. *et al.* (2012) Detection of recombination events in bacterial genomes from large population samples. *Nucleic acids …*. [Online] Available from: http://nar.oxfordjournals.org/content/40/1/e6.short.

Masterton, R.G. *et al.* (2008) Guidelines for the management of hospital-acquired pneumonia in the UK: report of the working party on hospital-acquired pneumonia of the British Society for Antimicrobial Chemotherapy. *The Journal of antimicrobial chemotherapy*. [Online] 62 (1), 5–34. Available from: doi:10.1093/jac/dkn162 [Accessed: 16 June 2015].

Mauldin, P.D. *et al.* (2010) Attributable hospital cost and length of stay associated with health care-associated infections caused by antibiotic-resistant gram-negative bacteria. *Antimicrobial agents and chemotherapy*. [Online] 54 (1), 109–115. Available from: doi:10.1128/AAC.01041-09 [Accessed: 19 June 2015].

McArthur, A.G. *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy*. [Online] 57 (7), 3348–3357. Available from: doi:10.1128/AAC.00419-13 [Accessed: 28 April 2014].

McClelland, M. *et al.* (2004) Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nature genetics*. [Online] 36 (12), 1268–1274. Available from: doi:10.1038/ng1470 [Accessed: 23 July 2015].

McCullough, J.L. & Maren, T.H. (1973) Inhibition of Dihydropteroate Synthetase from

Escherichia coli by Sulfones and Sulfonamides. *Antimicrobial Agents and Chemotherapy*. [Online] 3 (6), 665–669. Available from: doi:10.1128/AAC.3.6.665 [Accessed: 18 February 2015].

McFarland, L. V (2006) Meta-analysis of probiotics for the prevention of antibiotic associated diarrhea and the treatment of Clostridium difficile disease. *The American journal of gastroenterology*. [Online] 101 (4), 812–822. Available from: doi:10.1111/j.1572-0241.2006.00465.x [Accessed: 15 June 2015].

McNally, A. *et al.* (2013) The Evolutionary Path to Extraintestinal Pathogenic, Drug-Resistant Escherichia coli Is Marked by Drastic Reduction in Detectable Recombination within the Core Genome. *Genome biology and …*. [Online] 5 (4), 699–710. Available from: doi:10.1093/gbe/evt038 [Accessed: 20 May 2015].

Medini, D. *et al.* (2005) The microbial pan-genome. *Current Opinion in Genetics & Development*. [Online] 15 (6), 589–594. Available from: doi:10.1016/j.gde.2005.09.006.

Meeker, N.D. & Trede, N.S. (2008) Immunology and zebrafish: spawning new models of human disease. *Developmental and comparative immunology*. [Online] 32 (7), 745–757. Available from: doi:10.1016/j.dci.2007.11.011 [Accessed: 16 March 2015].

Melican, K. *et al.* (2011) Uropathogenic Escherichia coli P and Type 1 fimbriae act in synergy in a living host to facilitate renal colonization leading to nephron obstruction. *PLoS pathogens*. [Online] 7 (2), e1001298. Available from: doi:10.1371/journal.ppat.1001298 [Accessed: 16 April 2015].

Miller, S.I., Ernst, R.K. & Bader, M.W. (2005) LPS, TLR4 and infectious disease diversity. *Nature reviews. Microbiology*. [Online] 3 (1), 36–46. Available from: doi:10.1038/nrmicro1068 [Accessed: 9 August 2015].

Miskinyte, M. *et al.* (2013) The genetic basis of Escherichia coli pathoadaptation to macrophages. *PLoS pathogens*. [Online] 9 (12), e1003802. Available from: doi:10.1371/journal.ppat.1003802 [Accessed: 21 May 2015].

Miyata, S. *et al.* (2003) Use of the Galleria mellonella caterpillar as a model host to study the role of the type III secretion system in Pseudomonas aeruginosa pathogenesis. *Infection and immunity*. [Online] 71 (5), 2404–2413. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=153283&tool=pmcentrez&rendertype=abstract [Accessed: 31 July 2015].

Mogensen, T.H. (2009) Pathogen recognition and inflammatory signaling in innate immune defenses. *Clinical microbiology reviews*. [Online] 22 (2), 240–273, Table of Contents. Available from: doi:10.1128/CMR.00046-08 [Accessed: 14 July 2014].

Mora, A. *et al.* (2011) Characteristics of the Shiga-toxin-producing enteroaggregative Escherichia coli O104:H4 German outbreak strain and of STEC strains isolated in Spain. *International microbiology : the official journal of the Spanish Society for Microbiology*. [Online] 14 (3), 121–141. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22101411 [Accessed: 1 June 2015].

Morikane, K. (2012) Infection control in healthcare settings in Japan. *Journal of epidemiology / Japan Epidemiological Association*. [Online] 22 (2), 86–90. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3798586&tool=pmcentrez&rendertype=abstract [Accessed: 17 June 2015].

Morrison, D., Woodford, N. & Cookson, B. (1997) Enterococci as emerging pathogens of

humans. *Society for Applied Bacteriology symposium series*. [Online] 2689S – 99S. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9436321 [Accessed: 19 June 2015].

Morton, D.B. (1998) Ethical considerations in the use of animal models in infection. *Clinical Microbiology and Infection*. [Online] 4 (11), 613–614. Available from: doi:10.1111/j.1469-0691.1998.tb00341.x [Accessed: 11 May 2015].

Muniesa, M. *et al.* (2000) Characterization of a shiga toxin 2e-converting bacteriophage from an Escherichia coli strain of human origin. *Infection and immunity*. [Online] 68 (9), 4850–4855. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=101679&tool=pmcentrez&rendertype=abstract [Accessed: 27 April 2015].

Muñoz-López, M. & García-Pérez, J.L. (2010) DNA transposons: nature and applications in genomics. *Current genomics*. [Online] 11 (2), 115–128. Available from: doi:10.2174/138920210790886871 [Accessed: 3 February 2015].

Murray, P.R. *et al.* (2007) *Manual of clinical microbiology*. 9th edition. Patrick R Murray et al. (eds.). [Online]. Washington D.C., ASM Press. Available from: http://agris.fao.org/agris-search/search/display.do?f=2013/US/US2013012230001223.xml;US201300122335.

Nardini, M. & Dijkstra, B.W. (1999) α/β Hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology*. [Online] 9 (6), 732–737. Available from: doi:10.1016/S0959-440X(99)00037-8 [Accessed: 12 August 2015].

Nathwani, D. *et al.* (2014) Clinical and economic consequences of hospital-acquired resistant and multidrug-resistant Pseudomonas aeruginosa infections: a systematic review and meta-analysis. *Antimicrobial resistance and infection control*. [Online] 3 (1), 32. Available from: doi:10.1186/2047-2994-3-32 [Accessed: 26 May 2015].

Needham, A.J. *et al.* (2004) Drosophila melanogaster as a model host for Staphylococcus aureus infection. *Microbiology (Reading, England)*. [Online] 150 (Pt 7), 2347–2355. Available from: doi:10.1099/mic.0.27116-0 [Accessed: 19 April 2015].

Nessar, R. *et al.* (2012) Mycobacterium abscessus: a new antibiotic nightmare. *The Journal of antimicrobial chemotherapy*. [Online] 67 (4), 810–818. Available from: doi:10.1093/jac/dkr578 [Accessed: 6 September 2015].

Neuwirth, C. *et al.* (2006) VEB-1 in Achromobacter xylosoxidans from cystic fibrosis patient, France. *Emerging infectious diseases*. [Online] 12 (11), 1737–1739. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=17283625&retmode=ref&cmd=prlinks.

Nikaido, H. (2011) Structure and mechanism of RND-type multidrug efflux pumps. *Advances in enzymology and related areas of molecular biology*. [Online] 771–60. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3122131&tool=pmcentrez&rendertype=abstract [Accessed: 21 July 2015].

O'Sullivan, B.P. & Freedman, S.D. (2009) Cystic fibrosis. *Lancet*. [Online] 373 (9678), 1891–1904. Available from: doi:10.1016/S0140-6736(09)60327-5 [Accessed: 9 July 2014].

Ochiai, K. *et al.* (1959) Studies on inheritance of drug resistance between Shigella strains and E. coli strains. *Nippon Iji Shimpo*. 186134–46.

Ochman, H., Lawrence, J.G. & Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*. [Online] 405 (6784), 299–304. Available from: doi:10.1038/35012500 [Accessed: 30 January 2015].

Oh, D.-Y., Gaedicke, G. & Schreier, E. (2003) Viral agents of acute gastroenteritis in German children: prevalence and molecular diversity. *Journal of medical virology*. [Online] 71 (1), 82–93. Available from: doi:10.1002/jmv.10449 [Accessed: 15 June 2015].

Osborne, S.E. *et al.* (2009) Pathogenic adaptation of intracellular bacteria by rewiring a cis-regulatory input function. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 106 (10), 3982–3987. Available from: doi:10.1073/pnas.0811669106 [Accessed: 15 August 2015].

Otta, S. *et al.* (2014) Achromobacter xylosoxidans: a rare pathogen for community- acquired acute pancreatitis. *JMM Case Reports*. [Online] 1 (3), e001065. Available from: doi:10.1099/jmmcr.0.T00022 [Accessed: 19 December 2014].

Ou, J.T. & Anderson, T.F. (1970) Role of pili in bacterial conjugation. *Journal of bacteriology*. [Online] 102 (3), 648–654. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=247607&tool=pmcentrez&rendertype=abstract [Accessed: 17 August 2015].

Pallen, M.J., Beatson, S.A. & Bailey, C.M. (2005) Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion. *BMC microbiology*. [Online] 5 (1), 9. Available from: doi:10.1186/1471-2180-5-9 [Accessed: 3 August 2015].

Parashar, U.D. *et al.* (2006) Rotavirus and severe childhood diarrhea. *Emerging infectious diseases*. [Online] 12 (2), 304–306. Available from: doi:10.3201/eid1202.050006 [Accessed: 15 June 2015].

Park, P.-G. *et al.* (2012) GFP-tagged E. coli shows bacterial distribution in mouse organs: pathogen tracking using fluorescence signal. *Clinical and experimental vaccine research*. [Online] 1 (1), 83–87. Available from: doi:10.7774/cevr.2012.1.1.83 [Accessed: 10 August 2015].

Parkhill, J. *et al.* (2003) Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nature Genetics*. [Online] 35 (1), 32–40. Available from: doi:10.1038/ng1227.

Partridge, S.R. *et al.* (2009) Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiology Reviews*. [Online] 33 (4), 757–784. Available from: doi:10.1111/j.1574-6976.2009.00175.x.

Paulsen, I.T. *et al.* (1993) The 3' conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. *Antimicrobial agents and chemotherapy*. [Online] 37 (4), 761–768. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=187754&tool=pmcentrez&rendertype=abstract [Accessed: 28 July 2015].

Peleg, A.Y. *et al.* (2009) Galleria mellonella as a model system to study Acinetobacter baumannii pathogenesis and therapeutics. *Antimicrobial agents and chemotherapy*. [Online] 53 (6), 2605–2609. Available from: doi:10.1128/AAC.01533-08.

Pereira, R.H. V *et al.* (2011) Achromobacter xylosoxidans: Characterization of Strains in Brazilian Cystic Fibrosis Patients. *Journal of clinical microbiology*. [Online] 49 (10), 3649–3651. Available from: doi:10.1128/JCM.05283-11.

Phelan, P.E., Mellon, M.T. & Kim, C.H. (2005) Functional characterization of full-length TLR3, IRAK-4, and TRAF6 in zebrafish (Danio rerio). *Molecular immunology*. [Online] 42 (9), 1057–1071. Available from: doi:10.1016/j.molimm.2004.11.005 [Accessed: 6 March 2015].

Pien, F.D. & Higa, H.Y. (1978) Achromobacter xylosoxidans isolates in Hawaii. *Journal of*

clinical microbiology. [Online] 7 (2), 239–241. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=632351&retmode=ref&cmd=prlinks.

Di Pilato, V., Pollini, S. & Rossolini, G.M. (2014) Characterization of plasmid pAX22, encoding VIM-1 metallo-β-lactamase, reveals a new putative mechanism of In70 integron mobilization. *The Journal of antimicrobial chemotherapy*. [Online] 69 (1), 67–71. Available from: doi:10.1093/jac/dkt311.

Pili-Floury, S. *et al.* (2004) In vivo RNA interference analysis reveals an unexpected role for GNBP1 in the defense against Gram-positive bacterial infection in Drosophila adults. *The Journal of biological chemistry*. [Online] 279 (13), 12848–12853. Available from: doi:10.1074/jbc.M313324200 [Accessed: 27 April 2015].

Piotrowski, A., Luo, P. & Morrison, D.A. (2009) Competence for genetic transformation in Streptococcus pneumoniae: termination of activity of the alternative sigma factor ComX is independent of proteolysis of ComX and ComW. *Journal of bacteriology*. [Online] 191 (10), 3359–3366. Available from: doi:10.1128/JB.01750-08 [Accessed: 20 June 2015].

Piyakul, C., Tiyawisutsri, R. & Boonbumrung, K. (2012) Emergence of metallo-β-lactamase IMP-14 and VIM-2 in Pseudomonas aeruginosa clinical isolates from a tertiary-level hospital in Thailand. *Epidemiology and infection*. [Online] 140 (3), 539–541. Available from: doi:10.1017/S0950268811001294.

Plano, G. V & Straley, S.C. (1995) Mutations in yscC, yscD, and yscG prevent high-level expression and secretion of V antigen and Yops in Yersinia pestis. *Journal of bacteriology*. [Online] 177 (13), 3843–3854. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=177105&tool=pmcentrez&rendertype=abstract [Accessed: 3 August 2015].

Plowman, R. (2000) The socioeconomic burden of hospital acquired infection. *Eurosurveillance*. [Online] 4 (01). Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=4 [Accessed: 17 June 2015].

Pollack, M. (1984) The virulence of Pseudomonas aeruginosa. *Reviews of infectious diseases*. [Online] 6 (3), S617–S626. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6443764 [Accessed: 8 September 2015].

Powell, S. *et al.* (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic acids research*. [Online] 42 (Database issue), D231–D239. Available from: doi:10.1093/nar/gkt1253 [Accessed: 5 November 2014].

Pressley, M.E. *et al.* (2005) Pathogenesis and inflammatory response to Edwardsiella tarda infection in the zebrafish. *Developmental and comparative immunology*. [Online] 29 (6), 501–513. Available from: doi:10.1016/j.dci.2004.10.007 [Accessed: 27 April 2015].

Price, J.R. *et al.* (2014) Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of Staphylococcus aureus in an intensive care unit. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 58 (5), 609–618. Available from: doi:10.1093/cid/cit807 [Accessed: 3 June 2015].

Priyamvada, R. (2014) Pulmonary Infection Caused by Achromobacter xylosoxidans in a Patient with Carcinoma of Epiglottis: A Rare Case. *Journal of clinical and diagnostic research : JCDR*. 8 (5).

Public Health England (2014) *Surveillance of Surgical Site Infections in NHS Hospitals in*

England 2013/14. [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/386927/SSI_report_2013_14_final__3_.pdf [Accessed: 14 June 2015].

Puzniak, L.A. *et al.* (2004) A cost-benefit analysis of gown use in controlling vancomycin-resistant Enterococcus transmission: is it worth the price? *Infection control and hospital epidemiology*. [Online] 25 (5), 418–424. Available from: doi:10.1086/502416 [Accessed: 18 June 2015].

Qin, X. *et al.* (2012) Complete genome sequence of Enterococcus faecium strain TX16 and comparative genomic analysis of Enterococcus faecium genomes. *BMC microbiology*. [Online] 12 (1), 135. Available from: doi:10.1186/1471-2180-12-135.

Radford, A.D. *et al.* (2012) Application of next-generation sequencing technologies in virology. *The Journal of general virology*. [Online] 93 (Pt 9), 1853–1868. Available from: doi:10.1099/vir.0.043182-0 [Accessed: 9 July 2015].

Radics, J., Königsmaier, L. & Marlovits, T.C. (2014) Structure of a pathogenic type 3 secretion system in action. *Nature structural & molecular biology*. [Online] 21 (1), 82–87. Available from: doi:10.1038/nsmb.2722 [Accessed: 25 June 2015].

Ramirez, M.S. & Tolmasky, M.E. (2010) Aminoglycoside modifying enzymes. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy*. [Online] 13 (6), 151–171. Available from: doi:10.1016/j.drup.2010.08.003 [Accessed: 2 February 2015].

Rammelkamp, C.H. & Maxon, T. (1942) Resistance of Staphylococcus aureus to the Action of Penicillin. *Experimental Biology and Medicine*. [Online] 51 (3), 386–389. Available from: doi:10.3181/00379727-51-13986 [Accessed: 22 April 2015].

Rasko, D.A. *et al.* (2011) Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *The New England journal of medicine*. [Online] 365 (8), 709–717. Available from: doi:10.1056/NEJMoa1106920 [Accessed: 24 April 2015].

Rasko, D.A. *et al.* (2008) The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. *Journal of bacteriology*. [Online] 190 (20), 6881–6893. Available from: doi:10.1128/JB.00619-08 [Accessed: 5 May 2015].

Raso, T. *et al.* (2008) Achromobacter xylosoxidans respiratory tract infections in cystic fibrosis patients. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. [Online] 116 (9), 837–841. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19024606&retmode=ref&cmd=prlinks.

Reddy, S. *et al.* (2011) Epidemiology, antibiotic resistance trends and the cost of enteric fever in East London, 2005-2010. *Travel medicine and infectious disease*. [Online] 9 (4), 206–212. Available from: doi:10.1016/j.tmaid.2011.03.003 [Accessed: 27 July 2015].

Redgrave, L.S. *et al.* (2014) Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends in microbiology*. [Online] 22 (8), 438–445. Available from: doi:10.1016/j.tim.2014.04.007 [Accessed: 7 December 2014].

Reed, D. & Kemmerly, S.A. (2009) Infection control and prevention: a review of hospital-acquired infections and the economic implications. *The Ochsner journal*. [Online] 9 (1), 27–31. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3096239&tool=pmcentrez&rendertype=abstract [Accessed: 6 September 2015].

Reichman, D.E. & Greenberg, J.A. (2009) Reducing surgical site infections: a review. *Reviews in obstetrics & gynecology*. [Online] 2 (4), 212–221. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2812878&tool=pmcentrez&rendertype=abstract [Accessed: 26 April 2015].

Reina, J. *et al.* (1988) Nosocomial outbreak of Achromobacter xylosoxidans associated with a diagnostic contrast solution. *Journal of clinical pathology*. [Online] 41 (8), 920–921. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=3170779&retmode=ref&cmd=prlinks.

Relman, D.A. *et al.* (1992) Identification of the uncultured bacillus of Whipple's disease. *The New England journal of medicine*. 327293–301.

Reverdy, M.E. *et al.* (1984) Nosocomial colonization and infection by Achromobacter xylosoxidans. *Journal of clinical microbiology*. [Online] 19 (2), 140–143. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=6699141&retmode=ref&cmd=prlinks.

Ribeiro, F.J. *et al.* (2012) Finished bacterial genomes from shotgun sequence data. *Genome Research*. [Online] 22 (11), 2270–2277. Available from: doi:10.1101/gr.141515.112.

Riccio, M.L. *et al.* (2001) In70 of plasmid pAX22, a bla(VIM-1)-containing integron carrying a new aminoglycoside phosphotransferase gene cassette. *Antimicrobial agents and chemotherapy*. [Online] 45 (4), 1249–1253. Available from: doi:10.1128/AAC.45.4.1249-1253.2001.

Richardson, E.J. & Watson, M. (2013) The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*. [Online] 14 (1), 1–12. Available from: doi:10.1093/bib/bbs007.

Ripoll, F. *et al.* (2009) Non mycobacterial virulence genes in the genome of the emerging pathogen Mycobacterium abscessus. *PloS one*. [Online] 4 (6), e5660. Available from: doi:10.1371/journal.pone.0005660 [Accessed: 23 June 2015].

Roberts, R., Carneiro, M. & Schatz, M. (2013) The advantages of SMRT sequencing. *Genome Biology*. [Online] 14 (7), 405. Available from: doi:10.1186/gb-2013-14-7-405 [Accessed: 27 August 2015].

Rønne Hansen, C. *et al.* (2006) Chronic infection with Achromobacter xylosoxidans in cystic fibrosis patients; a retrospective case control study. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*. [Online] 5 (4), 245–251. Available from: doi:10.1016/j.jcf.2006.04.002.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)*. [Online] 19 (12), 1572–1574. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12912839 [Accessed: 27 November 2014].

Rouli, L. *et al.* (2014) Genomic analysis of three African strains of Bacillus anthracis demonstrates that they are part of the clonal expansion of an exclusively pathogenic bacterium. *New microbes and new infections*. [Online] 2 (6), 161–169. Available from: doi:10.1002/nmi2.62 [Accessed: 9 April 2015].

Rowe-Magnus, D.A. & Mazel, D. (2002) The role of integrons in antibiotic resistance gene capture. *International journal of medical microbiology : IJMM*. [Online] 292 (2), 115–125. Available from: doi:10.1078/1438-4221-00197.

Ruimy, R. *et al.* (2009) Comparisons between geographically diverse samples of carried Staphylococcus aureus. *Journal of bacteriology*. [Online] 191 (18), 5577–5583. Available from:

doi:10.1128/JB.00493-09 [Accessed: 20 August 2015].

Sabat, A.J. *et al.* (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance*. [Online] 18 (4). Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20380 [Accessed: 22 September 2015].

Saint, S. *et al.* (2008) Preventing hospital-acquired urinary tract infection in the United States: a national study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 46 (2), 243–250. Available from: doi:10.1086/524662 [Accessed: 15 June 2015].

Sánchez-Busó, L. *et al.* (2014) Recombination drives genome evolution in outbreak-related Legionella pneumophila isolates. *Nature genetics*. [Online] 46 (11), 1205–1211. Available from: doi:10.1038/ng.3114 [Accessed: 7 April 2015].

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 74 (12), 5463–5467. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=431765&tool=pmcentrez&rendertype=abstract [Accessed: 10 July 2014].

Sato, H. *et al.* (2003) The mechanism of action of the Pseudomonas aeruginosa-encoded type III cytotoxin, ExoU. *The EMBO journal*. [Online] 22 (12), 2959–2969. Available from: doi:10.1093/emboj/cdg290 [Accessed: 3 August 2015].

Sattar, S.A. *et al.* (1994) Interruption of rotavirus spread through chemical disinfection. *Infection control and hospital epidemiology*. [Online] 15 (12), 751–756. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7890922 [Accessed: 15 June 2015].

Sattar, S.A. *et al.* (1983) Rotavirus inactivation by chemical disinfectants and antiseptics used in hospitals. *Canadian journal of microbiology*. [Online] 29 (10), 1464–1469. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6318942 [Accessed: 15 June 2015].

Schattner, P., Brooks, A.N. & Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research*. [Online] 33 (Web Server issue), W686–W689. Available from: doi:10.1093/nar/gki366 [Accessed: 14 October 2014].

Schoen, C. *et al.* (2008) Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in Neisseria meningitidis. *Proceedings of the National Academy of Sciences*. [Online] 105 (9), 3473–3478. Available from: doi:10.1073/pnas.0800151105 [Accessed: 3 July 2015].

Schoeni, J.L. & Lee Wong, A.C. (2005) Bacillus cereus Food Poisoning and Its Toxins. *Journal of Food Protection*. [Online] 3 (13), 636–648. Available from: http://www.ingentaconnect.com/content/iafp/jfp/2005/00000068/00000003/art00035?token=00601f7fef478117409fd5539412f415d766b25247b552b6d7b7a516f253048296a7c2849266d656c829689d454d6723 [Accessed: 7 April 2015].

Schröttner, P. *et al.* (2014) Comparison of VITEK2, MALDI-TOF MS, and 16S rDNA sequencing for identification of Myroides odoratus and Myroides odoratimimus. *Diagnostic microbiology and infectious disease*. [Online] 79 (2), 155–159. Available from: doi:10.1016/j.diagmicrobio.2014.02.002 [Accessed: 27 May 2015].

Schwandner, R. *et al.* (1999) Peptidoglycan- and lipoteichoic acid-induced cell activation is

mediated by toll-like receptor 2. *The Journal of biological chemistry*. [Online] 274 (25), 17406–17409. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10364168 [Accessed: 11 May 2015].

Seed, K.D. & Dennis, J.J. (2008) Development of Galleria mellonella as an alternative infection model for the Burkholderia cepacia complex. *Infection and immunity*. [Online] 76 (3), 1267–1275. Available from: doi:10.1128/IAI.01249-07 [Accessed: 20 October 2014].

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*. [Online] 30 (14), 2068–2069. Available from: doi:10.1093/bioinformatics/btu153 [Accessed: 10 July 2014].

Sengstock, D.M. *et al.* (2010) Multidrug-resistant Acinetobacter baumannii: an emerging pathogen among older adults in community hospitals and nursing homes. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 50 (12), 1611–1616. Available from: doi:10.1086/652759 [Accessed: 11 June 2015].

Shallcross, L.J. & Davies, D.S.C. (2015) GP antibiotic overuse and microbial resistance. *The British journal of general practice : the journal of the Royal College of General Practitioners*. [Online] 65 (631), 61–62. Available from: doi:10.3399/bjgp15X683413 [Accessed: 27 July 2015].

Sheng, W.-H. *et al.* (2013) Distribution of extended-spectrum β-lactamases, AmpC β-lactamases, and carbapenemases among Enterobacteriaceae isolates causing intra-abdominal infections in the Asia-Pacific region: results of the study for Monitoring Antimicrobial Resistance Trends (SMA. *Antimicrobial agents and chemotherapy*. [Online] 57 (7), 2981–2988. Available from: doi:10.1128/AAC.00971-12.

Shigeta, S. *et al.* (1978) Cerebral ventriculitis associated with Achromobacter xylosoxidans. *Journal of clinical pathology*. [Online] 31 (2), 156–161. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=632360&retmode=ref&cmd=prlinks.

Shin, K.S. *et al.* (2005) Imipenem-resistant Achromobacter xylosoxidans carrying blaVIM-2-containing class 1 integron. *Diagnostic microbiology and infectious disease*. [Online] 53 (3), 215–220. Available from: doi:10.1016/j.diagmicrobio.2005.06.018.

Siarheyeva, A. & Sharom, F.J. (2009) The ABC transporter MsbA interacts with lipid A and amphipathic drugs at different sites. *The Biochemical journal*. [Online] 419 (2), 317–328. Available from: doi:10.1042/BJ20081364 [Accessed: 4 May 2015].

Smets, B.F. & Barkay, T. (2005) Horizontal gene transfer: perspectives at a crossroads of scientific disciplines. *Nature reviews. Microbiology*. [Online] 3 (9), 675–678. Available from: doi:10.1038/nrmicro1253 [Accessed: 26 August 2015].

Soares, S.C. *et al.* (2013) The pan-genome of the animal pathogen Corynebacterium pseudotuberculosis reveals differences in genome plasticity between the biovar ovis and equi strains. *PLOS ONE*. [Online] 8 (1), e53818. Available from: doi:10.1371/journal.pone.0053818 [Accessed: 20 August 2014].

Sofianou, D. *et al.* (2005) VIM-2 metallo-beta-lactamase in Achromobacter xylosoxidans in Europe. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 24 (12), 854–855. Available from: doi:10.1007/s10096-005-0054-2.

Somarajan, S.R. *et al.* (2014) CcpA is important for growth and virulence of Enterococcus faecium. *Infection and Immunity*. [Online] 82 (9), 3580–3587. Available from:

doi:10.1128/IAI.01911-14.

Song, L. *et al.* (2013) Genetic variability of mutans streptococci revealed by wide whole-genome sequencing. *BMC genomics*. [Online] 14430. Available from: doi:10.1186/1471-2164-14-430.

Sørum, H., Roberts, M.C. & Crosa, J.H. (1992) Identification and cloning of a tetracycline resistance gene from the fish pathogen Vibrio salmonicida. *Antimicrobial agents and chemotherapy*. [Online] 36 (3), 611–615. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=190565&tool=pmcentrez&rendertype=abstract [Accessed: 28 July 2015].

Spencer, R.C. (2003) Bacillus anthracis. *Journal of clinical pathology*. [Online] 56 (3), 182–187. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1769905&tool=pmcentrez&rendertype=abstract [Accessed: 7 April 2015].

Spicuzza, L. *et al.* (2009) Emerging pathogens in cystic fibrosis: ten years of follow-up in a cohort of patients. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 28 (2), 191–195. Available from: doi:10.1007/s10096-008-0605-4.

Spilker, T., Vandamme, P. & LiPuma, J.J. (2012) A multilocus sequence typing scheme implies population structure and reveals several putative novel Achromobacter species. *Journal of clinical microbiology*. [Online] 50 (9), 3010–3015. Available from: doi:10.1128/JCM.00814-12.

Stalder, T. *et al.* (2012) Integron involvement in environmental spread of antibiotic resistance. *Frontiers in microbiology*. [Online] 3119. Available from: doi:10.3389/fmicb.2012.00119.

Stokes, H.W. & Gillings, M.R. (2011) Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS microbiology reviews*. [Online] 35 (5), 790–819. Available from: doi:10.1111/j.1574-6976.2011.00273.x [Accessed: 14 July 2015].

Strnad, H. *et al.* (2011) Complete genome sequence of the haloaromatic acid-degrading bacterium Achromobacter xylosoxidans A8. *Journal of bacteriology*. [Online] 193 (3), 791–792. Available from: doi:10.1128/JB.01299-10.

Struelens, M.J. (1998) The epidemiology of antimicrobial resistance in hospital acquired infections: problems and possible solutions. *BMJ (Clinical research ed.)*. [Online] 317 (7159), 652–654. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1113836&tool=pmcentrez&rendertype=abstract [Accessed: 3 August 2015].

Su, Y.-C. *et al.* (2014) Comparative genomic analysis reveals distinct genotypic features of the emerging pathogen Haemophilus influenzae type f. *BMC genomics*. 1538.

Sumpradit, N. *et al.* (2012) Antibiotics Smart Use: a workable model for promoting the rational use of medicines in Thailand. *Bulletin of the World Health Organization*. [Online] 90 (12), 905. Available from: doi:10.2471/BLT.12.105445.

Sun, J., Deng, Z. & Yan, A. (2014) Bacterial multidrug efflux pumps: Mechanisms, physiology and pharmacological exploitations. *Biochemical and Biophysical Research Communications*. [Online] 453 (2), 254–267. Available from: doi:10.1016/j.bbrc.2014.05.090 [Accessed: 16 March 2015].

Swiatecka-Urban, A. *et al.* (2006) Pseudomonas aeruginosa inhibits endocytic recycling of CFTR in polarized human airway epithelial cells. *American journal of physiology. Cell physiology*.

[Online] 290 (3), C862–C872. Available from: doi:10.1152/ajpcell.00108.2005 [Accessed: 29 July 2015].

Tan, K.E. *et al.* (2012) Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: a bench-by-bench study for assessing the impact on ti. *Journal of clinical microbiology*. [Online] 50 (10), 3301–3308. Available from: doi:10.1128/JCM.01405-12 [Accessed: 9 May 2015].

Tan, M.W., Mahajan-Miklos, S. & Ausubel, F.M. (1999) Killing of Caenorhabditis elegans by Pseudomonas aeruginosa used to model mammalian bacterial pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 96 (2), 715–720. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9892699&retmode=ref&cmd=prlinks.

Tatro, E.T. *et al.* (2013) Antibody response to Achromobacter xylosoxidans during HIV infection is associated with lower CD4 levels and increased lymphocyte activation. *Clinical and vaccine immunology : CVI*. [Online] Available from: doi:10.1128/CVI.00553-13.

Tatusov, R.L. (1997) A Genomic Perspective on Protein Families. *Science*. [Online] 278 (5338), 631–637. Available from: doi:10.1126/science.278.5338.631 [Accessed: 11 July 2014].

Team, R.C. (2014) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. [Online] Available from: http://www.r-project.org/.

Tena, D. *et al.* (2005) Outbreak of long-term intravascular catheter-related bacteremia due to Achromobacter xylosoxidans subspecies xylosoxidans in a hemodialysis unit. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. [Online] 24 (11), 727–732. Available from: doi:10.1007/s10096-005-0028-4.

Tena, D. *et al.* (2014) Skin and soft tissue infection caused by Achromobacter xylosoxidans: report of 14 cases. *Scandinavian journal of infectious diseases*. [Online] 46 (2), 130–135. Available from: doi:10.3109/00365548.2013.857043.

Teo, J. *et al.* (2014) Comparative genomic analysis of malaria mosquito vector-associated novel pathogen Elizabethkingia anophelis. *Genome biology and evolution*. [Online] 6 (5), 1158–1165. Available from: doi:10.1093/gbe/evu094 [Accessed: 7 September 2015].

Tettelin, H. *et al.* (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*. [Online] 11 (5), 472–477. Available from: doi:10.1016/j.mib.2008.09.006.

Tettelin, H. *et al.* (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 102 (39), 13950–13955. Available from: doi:10.1073/pnas.0506758102.

Thamlikitkul, V. (1988) Antibiotic dispensing by drug store personnel in Bangkok, Thailand. *The Journal of antimicrobial chemotherapy*. [Online] 21 (1), 125–131. Available from: doi:10.1093/jac/21.1.125.

Thomas, C.M. & Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews. Microbiology*. [Online] 3 (9), 711–721. Available from: doi:10.1038/nrmicro1234 [Accessed: 11 July 2014].

Thomas, R.J. *et al.* (2013) Galleria mellonella as a model system to test the pharmacokinetics and

efficacy of antibiotics against Burkholderia pseudomallei. *International journal of antimicrobial agents*. [Online] 41 (4), 330–336. Available from: doi:10.1016/j.ijantimicag.2012.12.009.

Tlaskalová-Hogenová, H. *et al.* (2011) The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cellular & molecular immunology*. [Online] 8 (2), 110–120. Available from: doi:10.1038/cmi.2010.67 [Accessed: 10 August 2015].

Tomson, F.L. *et al.* (2005) Enteropathogenic Escherichia coli EspG disrupts microtubules and in conjunction with Orf3 enhances perturbation of the tight junction barrier. *Molecular microbiology*. [Online] 56 (2), 447–464. Available from: doi:10.1111/j.1365-2958.2005.04571.x [Accessed: 3 August 2015].

Torres, J.A., Villegas, M.V. & Quinn, J.P. (2007) Current concepts in antibiotic-resistant gram-negative bacteria. *Expert review of anti-infective therapy*. [Online] 5 (5), 833–843. Available from: doi:10.1586/14787210.5.5.833 [Accessed: 17 March 2015].

Touati, M. *et al.* (2013) Dissemination of a class I integron carrying VIM-2 carbapenemase in Pseudomonas aeruginosa clinical isolates from a hospital intensive care unit in Annaba, Algeria. *Antimicrobial agents and chemotherapy*. [Online] 57 (5), 2426–2427. Available from: doi:10.1128/AAC.00032-13.

Traglia, G.M. *et al.* (2012) Achromobacter xylosoxidans: an emerging pathogen carrying different elements involved in horizontal genetic transfer. *Current microbiology*. [Online] 65 (6), 673–678. Available from: doi:10.1007/s00284-012-0213-5 [Accessed: 16 June 2015].

Trancassini, M. *et al.* (2014) Outbreak of Achromobacter xylosoxidans in an Italian Cystic fibrosis center: genome variability, biofilm production, antibiotic resistance, and motility in isolated strains. *Frontiers in microbiology*. [Online] 5138. Available from: doi:10.3389/fmicb.2014.00138 [Accessed: 3 September 2014].

Tsukagoshi, N. & Aono, R. (2000) Entry into and release of solvents by Escherichia coli in an organic-aqueous two-liquid-phase system and substrate specificity of the AcrAB-TolC solvent-extruding pump. *Journal of bacteriology*. [Online] 182 (17), 4803–4810. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=111357&tool=pmcentrez&rendertype=abstract [Accessed: 16 February 2015].

Turton, J.F. *et al.* (2011) Identification of Achromobacter xylosoxidans by detection of the bla(OXA-114-like) gene intrinsic in this species. *Diagnostic microbiology and infectious disease*. [Online] 70 (3), 408–411. Available from: doi:10.1016/j.diagmicrobio.2011.02.007.

Tyson, G.H. *et al.* (2015) WGS accurately predicts antimicrobial resistance in Escherichia coli. *The Journal of antimicrobial chemotherapy*. [Online] 70 (10), 2763–2769. Available from: doi:10.1093/jac/dkv186 [Accessed: 12 September 2015].

Valiquette, L. *et al.* (2007) Impact of a reduction in the use of high-risk antibiotics on the course of an epidemic of Clostridium difficile-associated disease caused by the hypervirulent NAP1/027 strain. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. [Online] 45 Suppl 2 (Supplement_2), S112–S121. Available from: doi:10.1086/519258 [Accessed: 15 June 2015].

Vandamme, P. *et al.* (2013) Classification of Achromobacter genogroups 2, 5, 7 and 14 as Achromobacter insuavis sp. nov., Achromobacter aegrifaciens sp. nov., Achromobacter anxifer sp. nov. and Achromobacter dolens sp. nov., respectively. *Systematic and applied microbiology*. [Online] 36 (7), 474–482. Available from: doi:10.1016/j.syapm.2013.06.005.

Vann, W.F. *et al.* (2004) The NeuC Protein of Escherichia coli K1 Is a UDP N-Acetylglucosamine 2-Epimerase. *Journal of Bacteriology*. [Online] 186 (3), 706–712. Available from: doi:10.1128/JB.186.3.706-712.2004 [Accessed: 27 April 2015].

Vartivarian, S. *et al.* (1994) A changing pattern of susceptibility of Xanthomonas maltophilia to antimicrobial agents: implications for therapy. *Antimicrobial agents and chemotherapy*. [Online] 38 (3), 624–627. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=284510&tool=pmcentrez&rendertype=abstract [Accessed: 16 June 2015].

van Veen, S.Q., Claas, E.C.J. & Kuijper, E.J. (2010) High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. *Journal of clinical microbiology*. [Online] 48 (3), 900–907. Available from: doi:10.1128/JCM.02071-09 [Accessed: 24 February 2015].

Wagenlehner, F.M.E., Weidner, W. & Naber, K.G. (2005) Emergence of antibiotic resistance amongst hospital-acquired urinary tract infections and pharmacokinetic/pharmacodynamic considerations. *The Journal of hospital infection*. [Online] 60 (3), 191–200. Available from: doi:10.1016/j.jhin.2004.12.017 [Accessed: 15 June 2015].

Watanabe, A. *et al.* (2008) Risk factors associated with surgical site infection in upper and lower gastrointestinal surgery. *Surgery today*. [Online] 38 (5), 404–412. Available from: doi:10.1007/s00595-007-3637-y [Accessed: 14 June 2015].

Webber, M.A. (2002) The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy*. [Online] 51 (1), 9–11. Available from: doi:10.1093/jac/dkg050 [Accessed: 5 October 2014].

Weber, S.G. & Gold, H.S. (2003) Enterococcus: an emerging pathogen in hospitals. *Seminars in respiratory and critical care medicine*. [Online] 24 (1), 49–60. Available from: doi:10.1055/s-2003-37916 [Accessed: 16 September 2015].

Weisburg, W.G. *et al.* (1991) 16S ribosomal DNA amplification for phylogenetic study. *Journal of bacteriology*. [Online] 173 (2), 697–703. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=1987160&retmode=ref&cmd=prlinks.

Welsh, J. & McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic acids research*. [Online] 18 (24), 7213–7218. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=332855&tool=pmcentrez&rendertype=abstract [Accessed: 1 June 2015].

Wertheim, H.F.L. *et al.* (2009) Streptococcus suis: An Emerging Human Pathogen. *Clinical Infectious Diseases*. [Online] 48 (5), 617–625. Available from: doi:10.1086/596763 [Accessed: 5 April 2015].

Wilke, M. *et al.* (2011) Mouse models of cystic fibrosis: Phenotypic analysis and research applications. *Journal of Cystic Fibrosis*. [Online] 10S152–S171. Available from: doi:10.1016/S1569-1993(11)60020-9 [Accessed: 25 September 2015].

Willems, R.J.L. *et al.* (2005) Global Spread of Vancomycin- resistant Enterococcus faecium from Distinct Nosocomial Genetic Complex. *Emerging Infectious Diseases*. 11 (6), 821–828.

Willett, C.E., Cherry, J.J. & Steiner, L.A. (1997) Characterization and expression of the recombination activating genes (rag1 and rag2) of zebrafish. *Immunogenetics*. [Online] 45 (6), 394–404. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9089097 [Accessed: 27 April

2015].

Wilson, J.W. (2002) Mechanisms of bacterial pathogenicity. *Postgraduate Medical Journal*. [Online] 78 (918), 216–224. Available from: doi:10.1136/pmj.78.918.216 [Accessed: 23 March 2015].

Winham, S.J. *et al.* (2012) SNP interaction detection with Random Forests in high-dimensional genetic data. *BMC bioinformatics*. [Online] 13 (1), 164. Available from: doi:10.1186/1471-2105-13-164 [Accessed: 14 August 2015].

Winstanley, C. *et al.* (2009) Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Research*. [Online] 19 (1), 12–23. Available from: doi:10.1101/gr.086082.108.

Wittmann, J. *et al.* (2014) Isolation and Characterization of Numerous Novel Phages Targeting Diverse Strains of the Ubiquitous and Opportunistic Pathogen Achromobacter xylosoxidans. Stefan Bereswill (ed.). *PLOS ONE*. [Online] 9 (1), e86935. Available from: doi:10.1371/journal.pone.0086935 [Accessed: 5 September 2014].

Witz, G. & Stasiak, A. (2010) DNA supercoiling and its role in DNA decatenation and unknotting. *Nucleic acids research*. [Online] 38 (7), 2119–2133. Available from: doi:10.1093/nar/gkp1161 [Accessed: 24 August 2015].

Woebking, B. *et al.* (2005) Drug-lipid A interactions on the Escherichia coli ABC transporter MsbA. *Journal of bacteriology*. [Online] 187 (18), 6363–6369. Available from: doi:10.1128/JB.187.18.6363-6369.2005 [Accessed: 4 May 2015].

Woo, P.C. *et al.* (2001) Identification of Arcobacter cryaerophilus isolated from a traffic accident victim with bacteremia by 16S ribosomal RNA gene sequencing. *Diagnostic microbiology and infectious disease*. [Online] 40 (3), 125–127. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11502381 [Accessed: 30 May 2015].

Woo, P.C.Y. *et al.* (2008) Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. [Online] 14 (10), 908–934. Available from: doi:10.1111/j.1469-0691.2008.02070.x [Accessed: 28 March 2015].

Wright, G.D. (1999) Aminoglycoside-modifying enzymes. *Current Opinion in Microbiology*. [Online] 2 (5), 499–503. Available from: doi:10.1016/S1369-5274(99)00007-7 [Accessed: 8 September 2015].

Wuthiekanun, V. *et al.* (2009) Burkholderia Pseudomallei is genetically diverse in agricultural land in Northeast Thailand. *PLoS neglected tropical diseases*. [Online] 3 (8), e496. Available from: doi:10.1371/journal.pntd.0000496 [Accessed: 29 January 2015].

Yabuuchi, E. *et al.* (1998) Emendation of genus Achromobacter and Achromobacter xylosoxidans (Yabuuchi and Yano) and proposal of Achromobacter ruhlandii (Packer and Vishniac) comb. nov., Achromobacter piechaudii (Kiredjian et al.) comb. nov., and Achromobacter xylosoxidans subsp. de. *Microbiology and immunology*. [Online] 42 (6), 429–438. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9688077&retmode=ref&cmd=prlinks.

Yabuuchi, E. & Oyama, A. (1971) Achromobacter xylosoxidans n. sp. from human ear discharge.

Japanese journal of microbiology. [Online] 15 (5), 477–481. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=5316576&retmode=ref&cmd=prlinks.

Yabuuchi, E. & Yano, I. (1981) Achromobacter gen. nov. and Achromobacter xylosoxidans (ex Yabuuchi and Ohyama 1971) nom. rev. *International journal of systematic bacteriology*. [Online] 31 (4), 477–478. Available from: doi:10.1099/00207713-31-4-477.

Yamamoto, M. *et al.* (2012) Molecular characterization of IMP-type metallo-β-lactamases among multidrug-resistant Achromobacter xylosoxidans. *The Journal of antimicrobial chemotherapy*. [Online] 67 (9), 2110–2113. Available from: doi:10.1093/jac/dks179.

You, E. *et al.* (2014) Reduction in the incidence of hospital-acquired Clostridium difficile infection through infection control interventions other than the restriction of antimicrobial use. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*. [Online] 229–10. Available from: doi:10.1016/j.ijid.2014.01.011 [Accessed: 15 June 2015].

Ypma, R.J.F., van Ballegooijen, W.M. & Wallinga, J. (2013) Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics*. [Online] 195 (3), 1055–1062. Available from: doi:10.1534/genetics.113.154856 [Accessed: 22 June 2015].

Zarb, P. *et al.* (2012) The European Centre for Disease Prevention and Control (ECDC) pilot point prevalence survey of healthcare-associated infections and antimicrobial use. *Eurosurveillance*. [Online] 17 (46). Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20316 [Accessed: 12 June 2015].

Zerbino, D.R. & Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. [Online] 18 (5), 821–829. Available from: doi:10.1101/gr.074492.107.

Zhang, C. *et al.* (2007) Aerobic degradation of bisphenol A by Achromobacter xylosoxidans strain B-16 isolated from compost leachate of municipal solid waste. *Chemosphere*. [Online] 68 (1), 181–190. Available from: doi:10.1016/j.chemosphere.2006.12.012.

Zhao, W.-H. & Hu, Z.-Q. (2011) IMP-type metallo-β-lactamases in Gram-negative bacilli: distribution, phylogeny, and association with integrons. *Critical Reviews in Microbiology*. [Online] 37 (3), 214–226. Available from: doi:10.3109/1040841X.2011.559944.

Zhou, J., Bowler, L.D. & Spratt, B.G. (1997) Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of Neisseria meningitidis and commensal Neisseria species. *Molecular microbiology*. [Online] 23 (4), 799–812. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9157250 [Accessed: 7 April 2015].

Zhu, Y. *et al.* (2014) Distribution and molecular profiling of class 1 integrons in MDR Acinetobacter baumannii isolates and whole genome-based analysis of antibiotic resistance mechanisms in a representative strain. *Microbiological research*. [Online] 169 (11), 811–816. Available from: doi:10.1016/j.micres.2014.04.002 [Accessed: 22 June 2015].

Zimlichman, E. *et al.* (2013) Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. *JAMA internal medicine*. [Online] 173 (22), 2039–2046. Available from: doi:10.1001/jamainternmed.2013.9763 [Accessed: 17 June 2015].

Zou, W. *et al.* (2013) Meta-analysis of pulsed-field gel electrophoresis fingerprints based on a constructed Salmonella database. *PloS one*. [Online] 8 (3), e59224. Available from: doi:10.1371/journal.pone.0059224 [Accessed: 1 June 2015].

Appendices

List of Appendix Figures

List of Appendix Tables

Appendix 1 Scripts and Command lines used in this study

1.1 Command line used to perform reciprocal BLAST and to parse functional assignment via eggNOG. A script is taken from FAS Center for System Biology, Harvard University (http://archive.sysbio.harvard.edu/csb/resources/computational/scriptome/UNIX/Protocols/Sequences.html) and edited for in-house use.

blastall -p blastp -i <A file containing amino acid sequence of the strain> -d /pub16/pisut/PROJECTS/eggNOG/DB/aln_db/bactNOG/bactNOG.fasta -e 1e-10 -b 3 -v 3 -a 16 -m 8 -o A_B

blastall -p blastp -d < A file containing amino acid sequence of the strain > -i /pub16/pisut/PROJECTS/eggNOG/DB/aln_db/bactNOG/bactNOG.fasta -e 1e-10 -b 3 -v 3 -a 16 -m 8 -o B_A

perl -e '$name_col=0; $score_col=11; while(<>) {s/\r?\n//; @F=split /\t/, $_; ($n, $s) = @F[$name_col, $score_col]; if (! exists($max{$n})) {push @names, $n}; if (! exists($max{$n}) || $s > $max{$n}) {$max{$n} = $s; $best{$n} = ()}; if ($s == $max{$n}) {$best{$n} .= "$_\n"};} for $n (@names) {print $best{$n}}' A_B > A_B.best

perl -e '$name_col=0; $score_col=11; while(<>) {s/\r?\n//; @F=split /\t/, $_; ($n, $s) = @F[$name_col, $score_col]; if (! exists($max{$n})) {push @names, $n}; if (! exists($max{$n}) || $s > $max{$n}) {$max{$n} = $s; $best{$n} = ()}; if ($s == $max{$n}) {$best{$n} .= "$_\n"};} for $n (@names) {print $best{$n}}' B_A > B_A.best

```
perl -e '$col1=1; $col2=0;' -e '($f1,$f2)=@ARGV; open(F1,$f1); while (<F1>) {s/\r?\n//; @F=split /\t/, $_; $line1{$F[$col1]} .= "$_\n"} open(F2,$f2); while (<F2>) {s/\r?\n//;@F=split /\t/, $_; if ($x = $line1{$F[$col2]}) {$x =~ s/\n/\t$_\n/g; print $x}}' A_B.best B_A.best > A_B_A
```

```
perl -e '$colm=0; $coln=13; $count=0; while(<>) {s/\r?\n//; @F=split /\t/, $_; if ($F[$colm] eq $F[$coln]) {print "$_\n"; $count++}} warn "\nChose $count lines out of $. where column $colm had same text as column $coln\n\n";' A_B_A > A_B_A.recip
```

```
perl -e '@cols=(0, 1, 10, 11); while(<>) {s/\r?\n//; @F=split /\t/, $_; print join("\t", @F[@cols]), "\n"} warn "\nJoined columns ", join(", ", @cols), " for $. lines\n\n"' A_B_A.recip > RBHB.out
```

```
cat RBHB.out | awk '{print $2}' > use.bbh.txt
```

```
perl /pub16/pisut/PROJECTS/perl_scripts/str_match.pl use.bbh.txt /pub16/pisut/PROJECTS/eggNOG/DB/bactNOG.members.adjusted.txt  > eggNOG_prot.txt
```

```
cat eggNOG_prot.txt | awk '{print $2}' > use.eggNOG_prot.txt
```

```
perl /pub16/pisut/PROJECTS/perl_scripts/str_match.pl use.eggNOG_prot.txt /pub16/pisut/PROJECTS/eggNOG/DB/bactNOG.description.txt > descrip.txt
```

```
perl /pub16/pisut/PROJECTS/perl_scripts/str_match.pl use.eggNOG_prot.txt /pub16/pisut/PROJECTS/eggNOG/DB/bactNOG.funccat.txt > COG.txt
```

1.2 Commands used to parse OrthoMCL's output to generate a binary table presenting the presence/absence of genes in bacterial strains. Perl's and Pyhton's scripts are written with a help from Thomas Craig.

1) Create fasta files with every protein used in clustering

for i in *.gene.faa; do perl ~/PROJECTS/orthomcl/from_jen/append_header_SA.pl $i > $i.append; done #append headers of fasta files to look like '>geneID(filename)' (which are similar to orthoMCL output)

cat *.append > ALLfiles.append # concatenating all appended files together (make a new file containing all files)

2) Generate a binary table using appended file and OrthoMCL's output

./table_alt_cluster.py ALLfiles.append all_orthomcl.out > table.out

1.3 Methods for Pan-genome and Core genome extrapolation

1) Uploading code into R

load("/Users/BEER/Desktop/AXresearch/Pan-Core_Genome/For_Jen_Redesigned_code")

2) Uploading the binary table generated using scripts in Appendix 1.2

data<-as.matrix(read.table("binary table"))

3) Run 'Wrapper' with 1,000 permutations (a command from the previously uploaded code)

Wrapper(data,1000) -> all_out_data

4) Write CORE genes to a file

all_out_data[,2,] -> CORE
write.csv(CORE,"core.out ")

5) Write PAN genes to a file

all_out_data[,3,] -> PAN
write.csv(PAN,"pan.out")

6) Fitting core-genome and pan-genome to exponential model and power model, respectively

Appendix 2 Comparative Genomic analysis

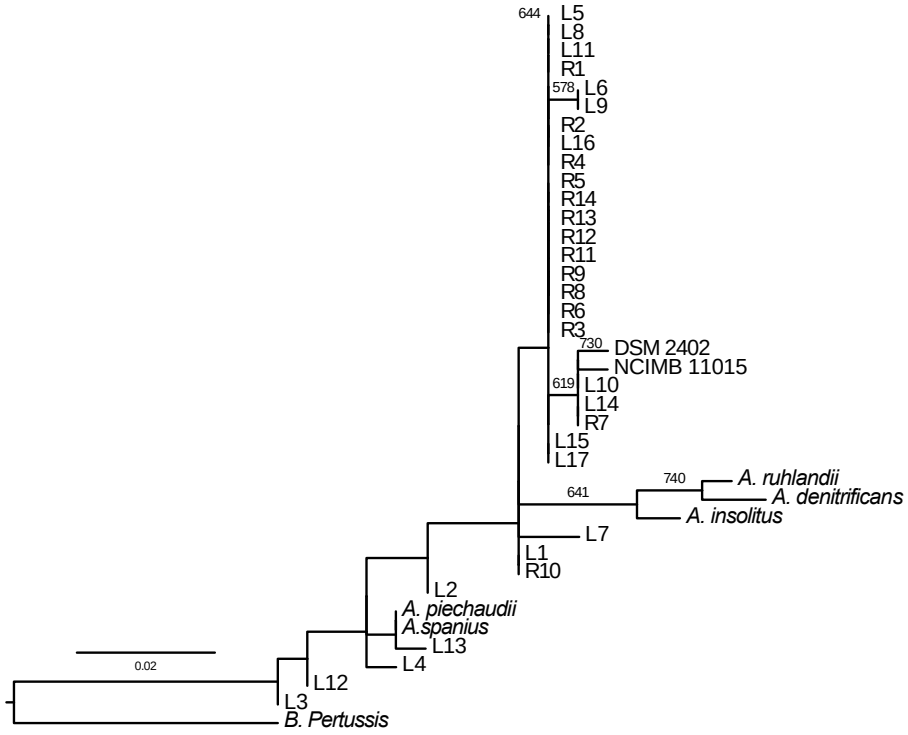


Figure A2.1: Phylogenetic tree based on maximum-likelihood analysis of 16S rDNA gene sequences of *Achromobacter* isolates. *B. pertussis* was used as an outgroup. The tree is built on 1,000 bootstrap repeats. R1-R14 are Thai clinical isolates; L1-L17 are British clinical isolates; DSM2402 stands for *A. xylosoxidans* DSM 2402; and AX11015 for *A. xylosoxidans* NCIMB 11015. Bootstrap values >50% are indicated.

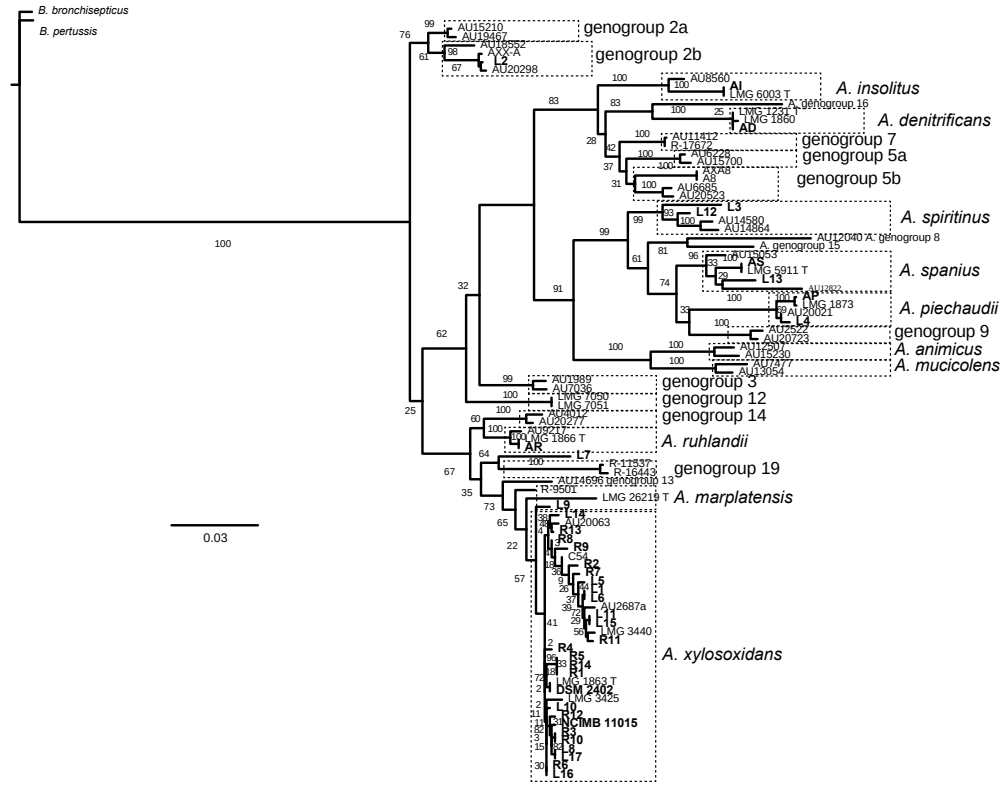


Figure A2.2: Phylogenetic tree of the *Achromobacter* species based on the analysis of MLST scheme. The tree is constructed by Maximum-likelihood with 100 bootstrap repeats. Strains used in this study are presented in bold. *Bordetella bronchisepticus* and *Bordetella pertussis* are used as outgroup.

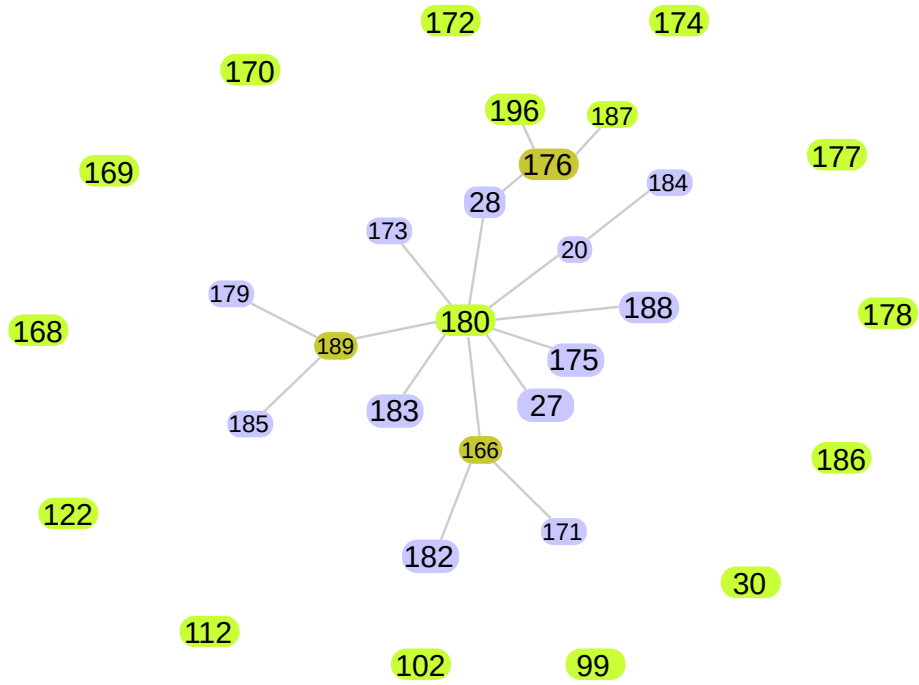


Figure A2.3: A population snapshot based on *Achromobacter* isolates used in this study. This shows the differentiation between *A. xylosoxidans* (forming a clonal complex) and non-*xylosoxidans* species (singletons).

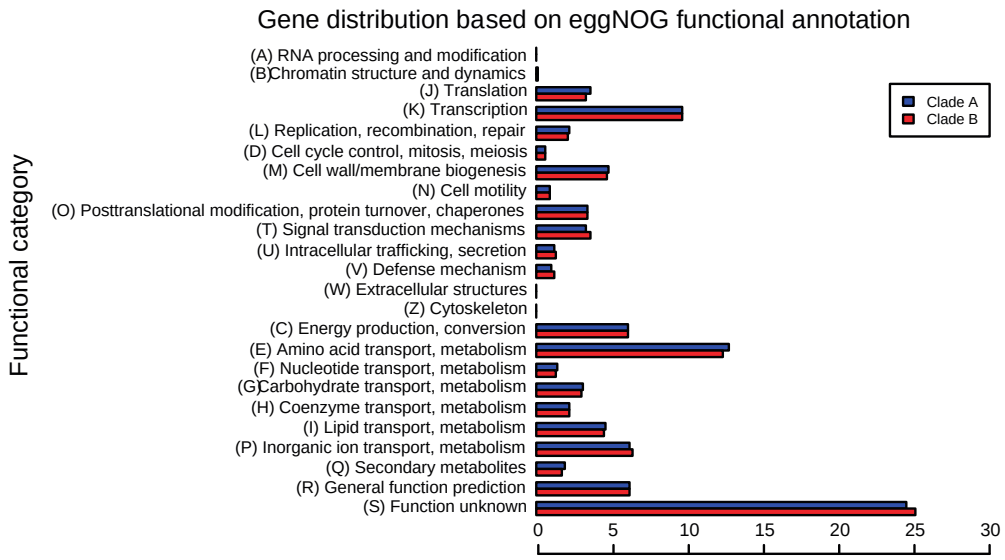


Figure A2.4: Gene distribution on core genome of *A. xylosoxidans* presented in clade A and clade B on core genome phylogeny (Figure 4.5 and 4.6). X-axis indicates the percentage of gene with respect to core genome. Y-axis indicates functional category based on eggNOG database.

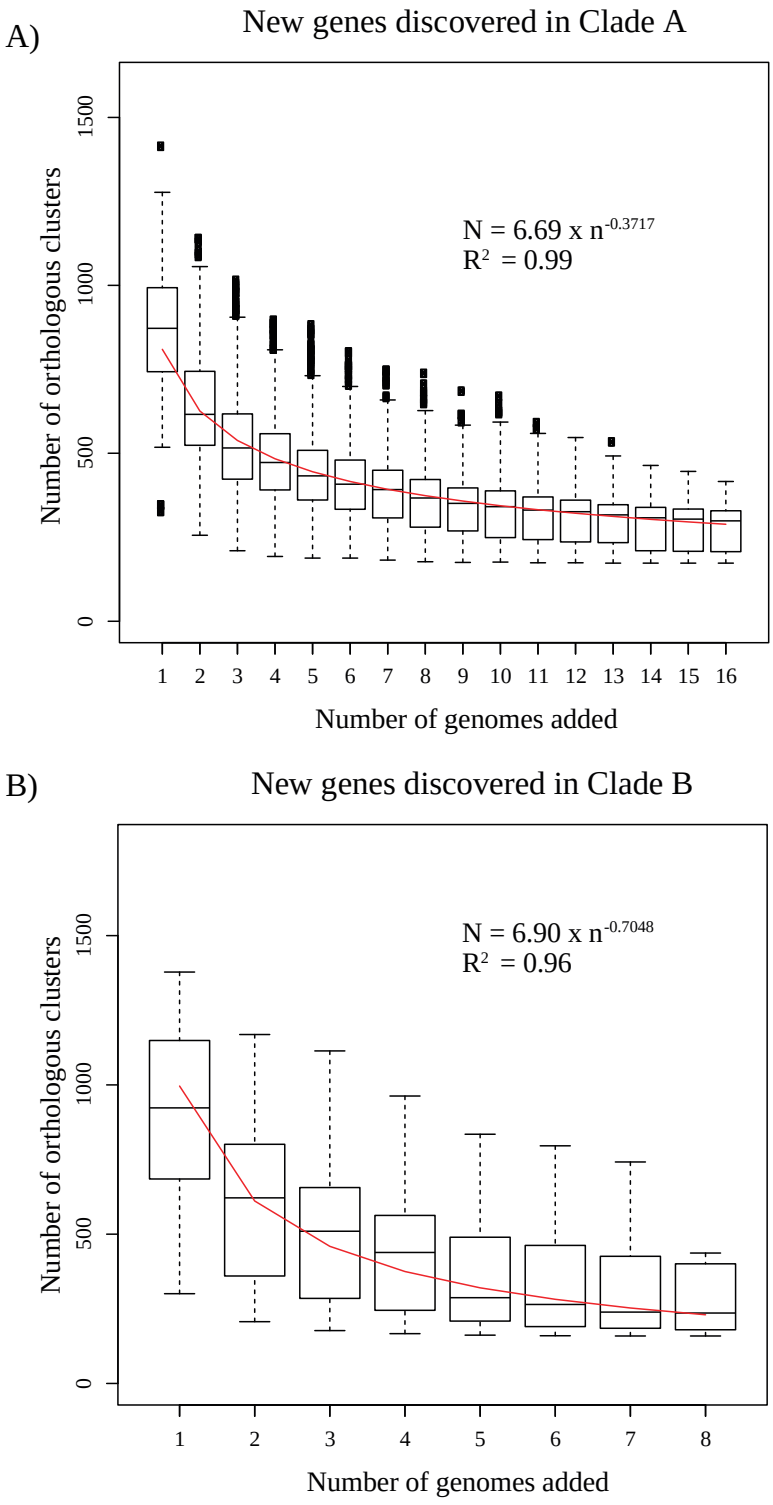


Figure A2.5: Regression curve for new gene discovered with additional genomes in clade I (A) and clade II (B) of *A. xylosoxidans*. Extrapolated mathematical model is reported.

Table A2.1: **The mapping percentage of sequence reads of bacterial strains used in this study to the genome of *A. xylosoxidans* NH44784-1996 (the second column) and to strain NCIMB 11015 (the third column).** The genome mapping is performed using BWA version 0.5.9-r16 and mapping statistic is obtained using SAMtools version 0.1.18-r580.

Strain	Percentage of reads mapped to NH44784-1996 genome	Percentage of reads mapped to NCIMB 11015 genome
NCIMB 11015	85.56%	99.68%*
DSM 2402	84.66%	83.31%
L1	96.12%	83.72%
L5	88.80%	86.87%
L8	88.48%	82.92%
L10	84.72%	80.89%
L11	85.61%	80.80%
L14	88.95%	85.33%
L15	80.76%	76.10%
L16	86.17%	83.80%
L17	89.66%	82.87%
R1	84.84%	79.63%
R2	89.68%	86.92%
R3	85.62%	82.78%
R4	85.22%	85.37%
R5	85.24%	80.48%
R6	88.57%	84.45%
R7	86.51%	81.47%
R8	83.17%	80.55%
R9	89.09%	87.23%
R10	87.88%	85.05%
R11	91.72%	89.32%
R12	91.26%	89.15%
R13	87.71%	85.70%
R14	85.91%	81.17%
Mean	87.28%	84.22%

***** Self-mapping of Illumina's reads to sequence obtained from PacBio RS II platform

Table A2.1: The mapping percentage of sequence reads of bacterial strains used in this study to the genome of *A. xylosoxidans* NH44784-1996 (the second column) and to strain NCIMB 11015 (the third column). The genome mapping is performed using BWA version 0.5.9-r16 and mapping statistic is obtained using SAMtools version 0.1.18-r580 (Continued)

Strain	Percentage of reads mapped to NH44784-1996 genome	Percentage of reads mapped to NCIMB 11015 genome
A. denitrificans	13.25%	13.33%
A. insolitus	9.17%	8.97%
A. piechaudii	13.25%	13.61%
A. ruhlandii	53.80%	52.17%
A. spanius	8.80%	7.11%
L2	52.78%	53%
L3	11.96%	10.43%
L4	10.76%	9.22%
L6	77.27%	74.19%
L7	41.75%	40.81%
L9	79.57%	78.07%
L12	10.54%	9.43%
L13	8.35%	8.07%
Mean	30.10%	29.11%

Table A2.2: The number of genes unique to each *A. xylosoxidans* isolate used in this study.

Isolates	No. of genes
NCIMB 11015	167
DSM 2402	190
NH44784-1996	141
L1	101
L5	140
L8	100
L10	174
L11	82
L14	115
L15	96
L16	129
L17	115
R1	89
R2	97
R3	91
R4	145
R5	87
R6	141
R7	135
R8	180
R9	99
R10	91
R11	99
R12	99
R13	105
R14	85
Mean	118.96
Median	103
S.D.	32.05

Table A2.3: The functional annotation of core orthologs and accessory orthologs of 26 genomes of *A. xylosoxidans* with respect to Cluster of Orthologous Group (COG) implemented using eggNOG database

COGs function	%Core orthologs	%Accessory orthologs
Information storage and processing		
(A) RNA processing and modification	0.03%	0.00%
(B) Chromatin structure and dynamics	0.13%	0.00%
(J) Translation	3.69%	0.74%
(K) Transcription	9.64%	6.34%
(L) Replication, recombination, repair	2.21%	3.34%
Cellular processes		
(D) Cell cycle control, mitosis, meiosis	0.61%	0.34%
(M) Cell wall/membrane biogenesis	4.91%	2.70%
(N) Cell motility	0.90%	0.50%
(O) Posttranslational modification, protein turnover, chaperones	3.53%	1.29%
(T) Signal transduction mechanisms	3.40%	1.89%
(U) Intracellular trafficking, secretion	1.25%	1.31%
(V) Defense mechanisms	1.06%	1.15%
(W) Extracellular structures	0%	0.00%
(Z) Cytoskeleton	0%	0.00%
Metabolism		
(C) Energy production, conversion	5.95%	3.80%
(E) Amino acid transport, metabolism	12.57%	5.41%
(F) Nucleotide transport, metabolism	1.41%	0.40%
(G) Carbohydrate transport, metabolism	3.19%	1.37%
(H) Coenzyme transport, metabolism	2.28%	0.64%
(I) Lipid transport, metabolism	4.60%	1.87%
(P) Inorganic ion transport, metabolism	6.16%	4.16%
(Q) Secondary metabolites	1.78%	1.15%
Poorly characterised		
(R) General function prediction	6.19%	4.06%
(S) Function unknown	24.51%	57.53%

Appendix 3 Antibiotic resistance determination

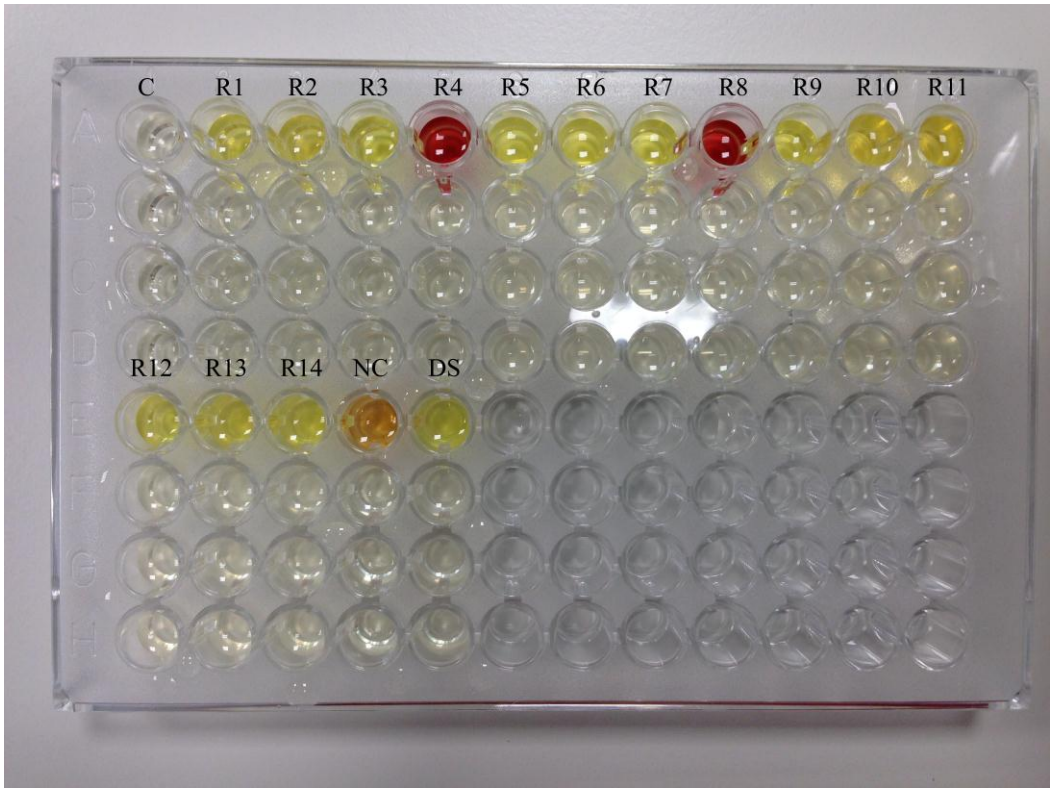


Figure A3.1: The detection of Beta-lactamase activity in *A. xylosoxidans* using Nitrocefin assay. The significant change of visible colour was observed in strain R4, R8 and NCIMB 11015 after incubation for 25 minutes, C stands for negative control; R1-R14 for Thai strains; NC for NCMIB11015; and DS for DSM2402

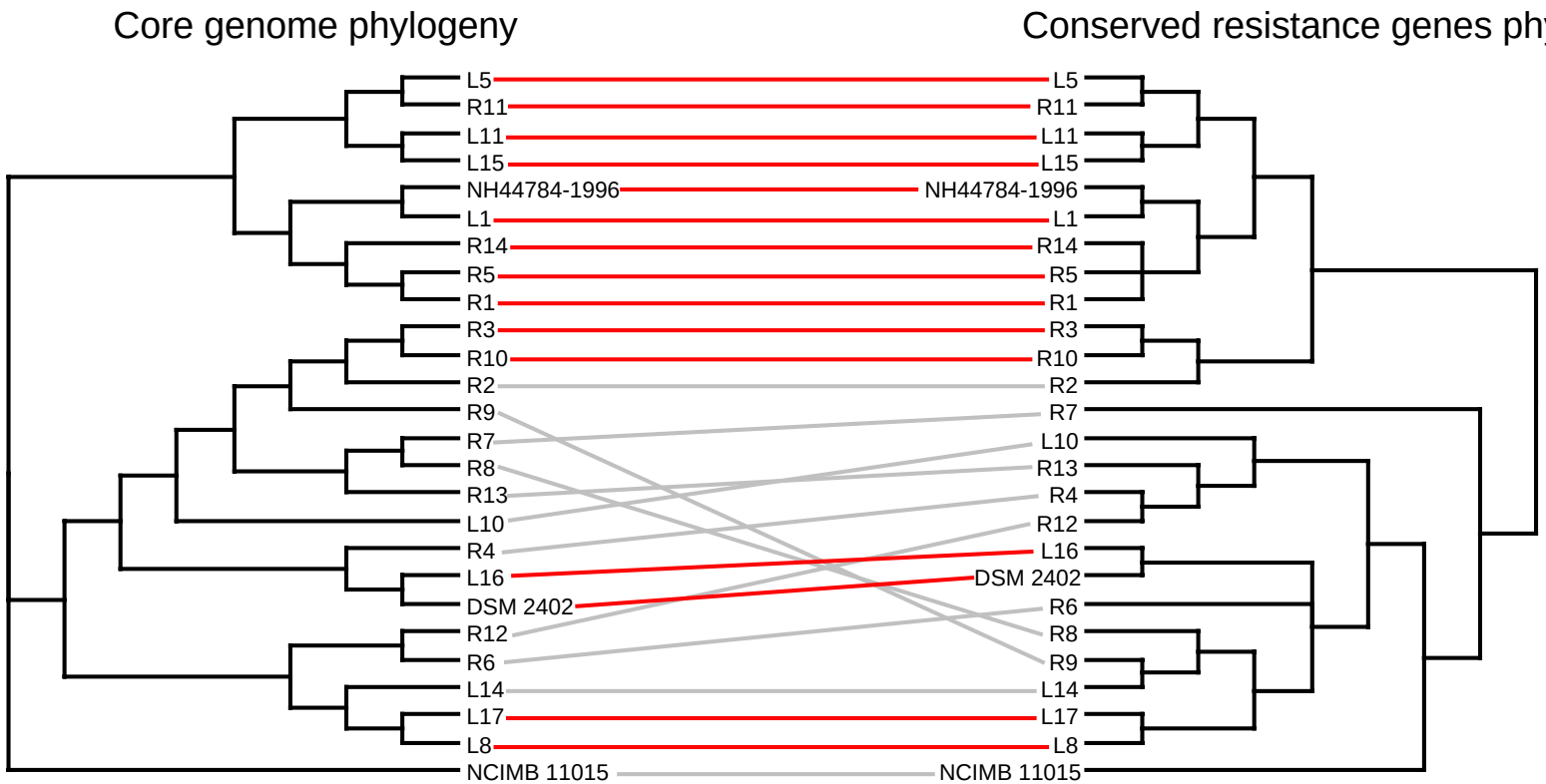


Figure A3.2: A tanglegram tree to show the association between the core genome phylogenetic tree (left panel) and an intrinsic antibiotic resistance phylogenetic tree (right panel). Red lines indicate consistent relationships between isolates in the phylogenetic trees based on core genes and conserved antibiotic resistance genes. NH44784-1996 stands for *A. xylosoxidans* NH44784-1996, NCIMB 11015 for *A. xylosoxidans* NCIMB 11015; DSM 2402 for *A. xylosoxidans* DSM 2402; L-strains are Liverpool's clinical isolates; R-strains are Thai clinical isolates.

Figure A3.3: Multiple sequence alignment of CmeABC genes and AxyXY-OprZ

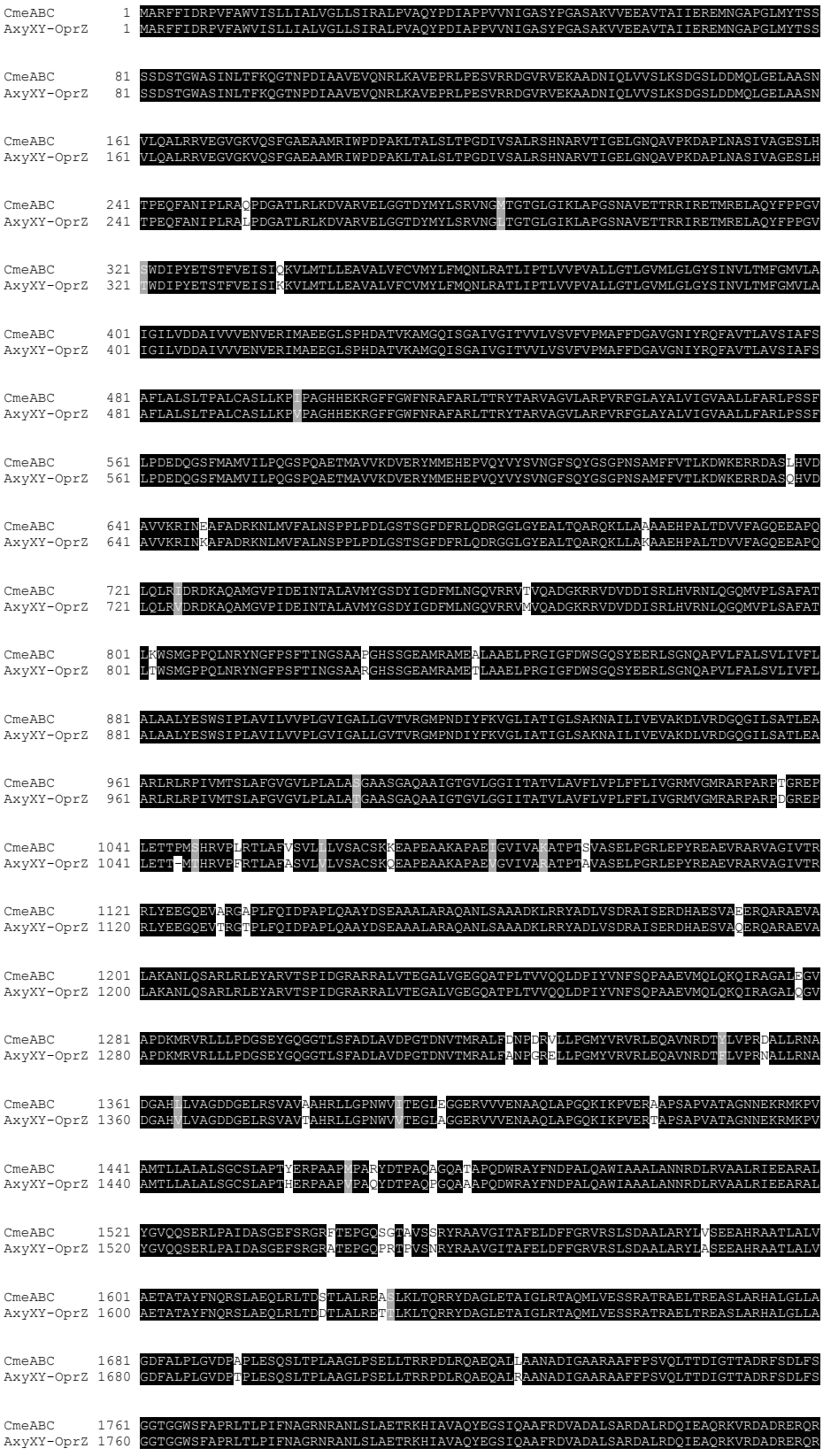
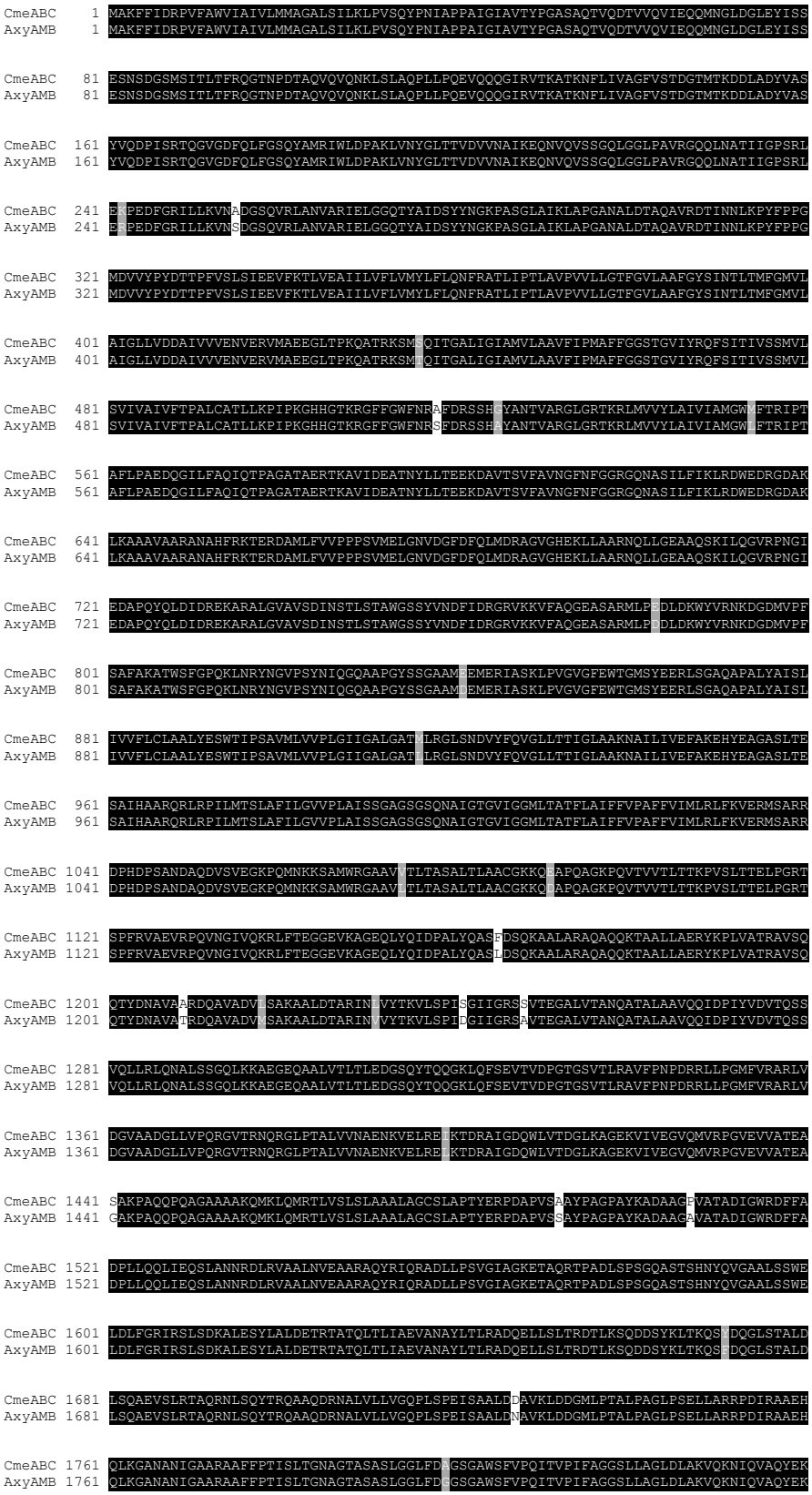


Figure A3.4: Multiple sequence alignment of CmeABC genes and AxyABM



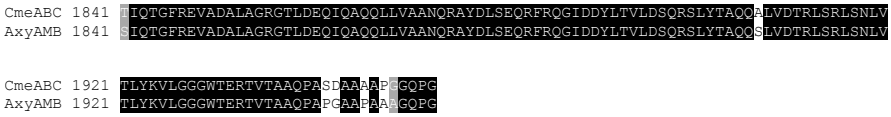
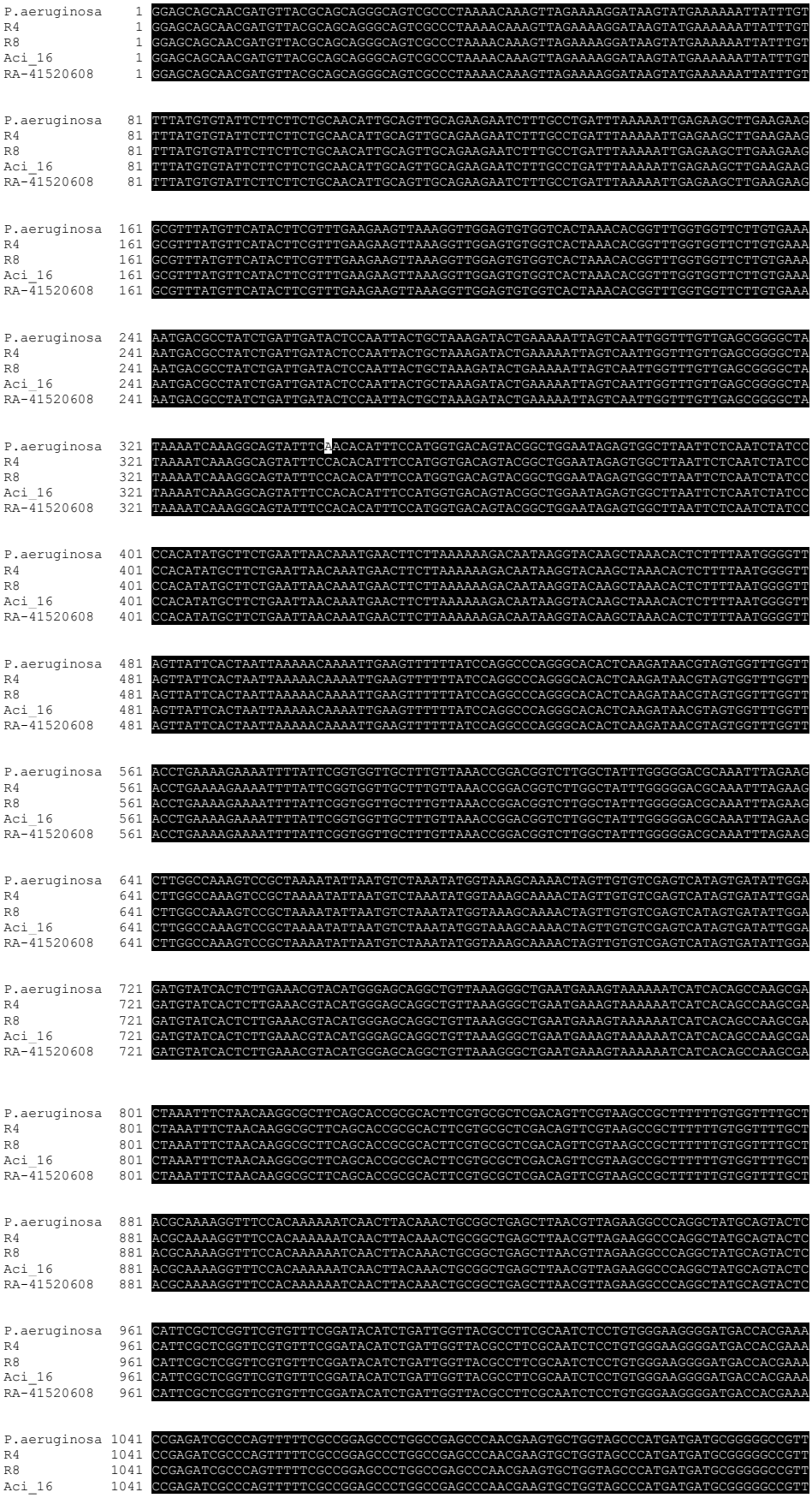

CmeABC 1841 IQTGFREVADALAGRGTLDEQIQAQQLLVAANQRAYDLSEQRFRQGIDDYLTVLDSQRSLYTAQQALVDTRLSRLSNLV
AxyAMB 1841 IQTGFREVADALAGRGTLDEQIQAQQLLVAANQRAYDLSEQRFRQGIDDYLTVLDSQRSLYTAQQSLVDTRLSRLSNLV
CmeABC 1921 TLYKVLGGGWTERTVTAAQPASDAAAAPGGQPG
AxyAMB 1921 TLYKVLGGGWTERTVTAAQPAPAAPAAGQPG

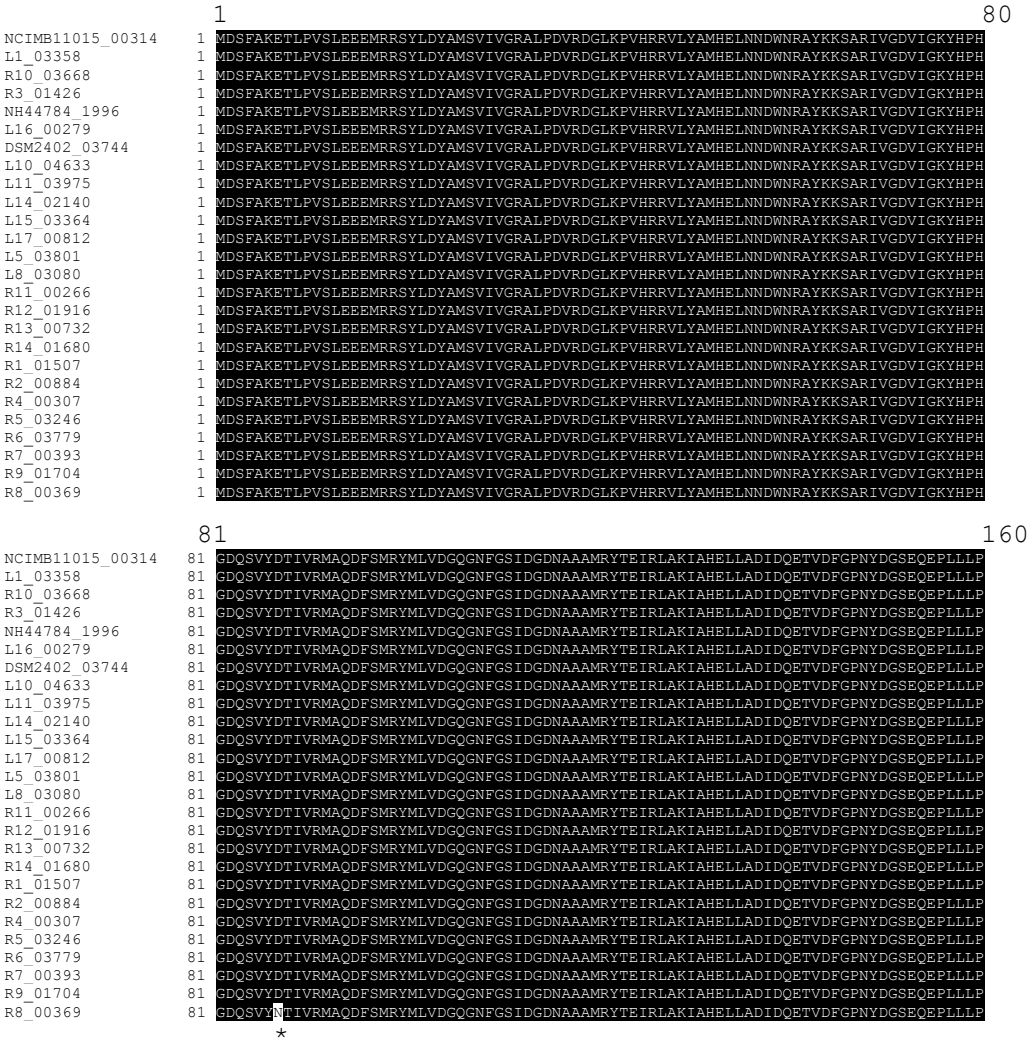
Figure A3.5: Multiple sequence alignment of gene cassettes with the integron

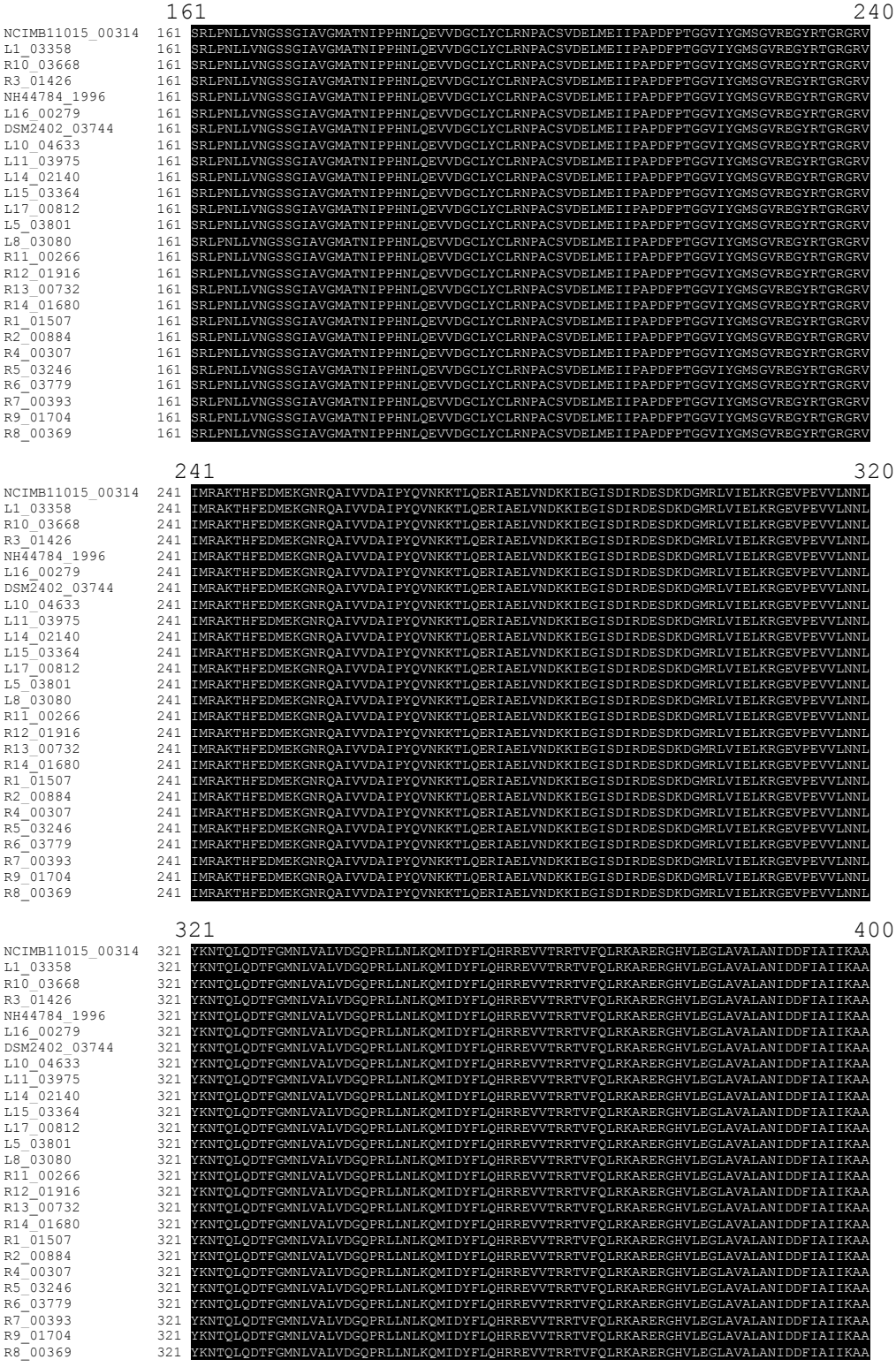


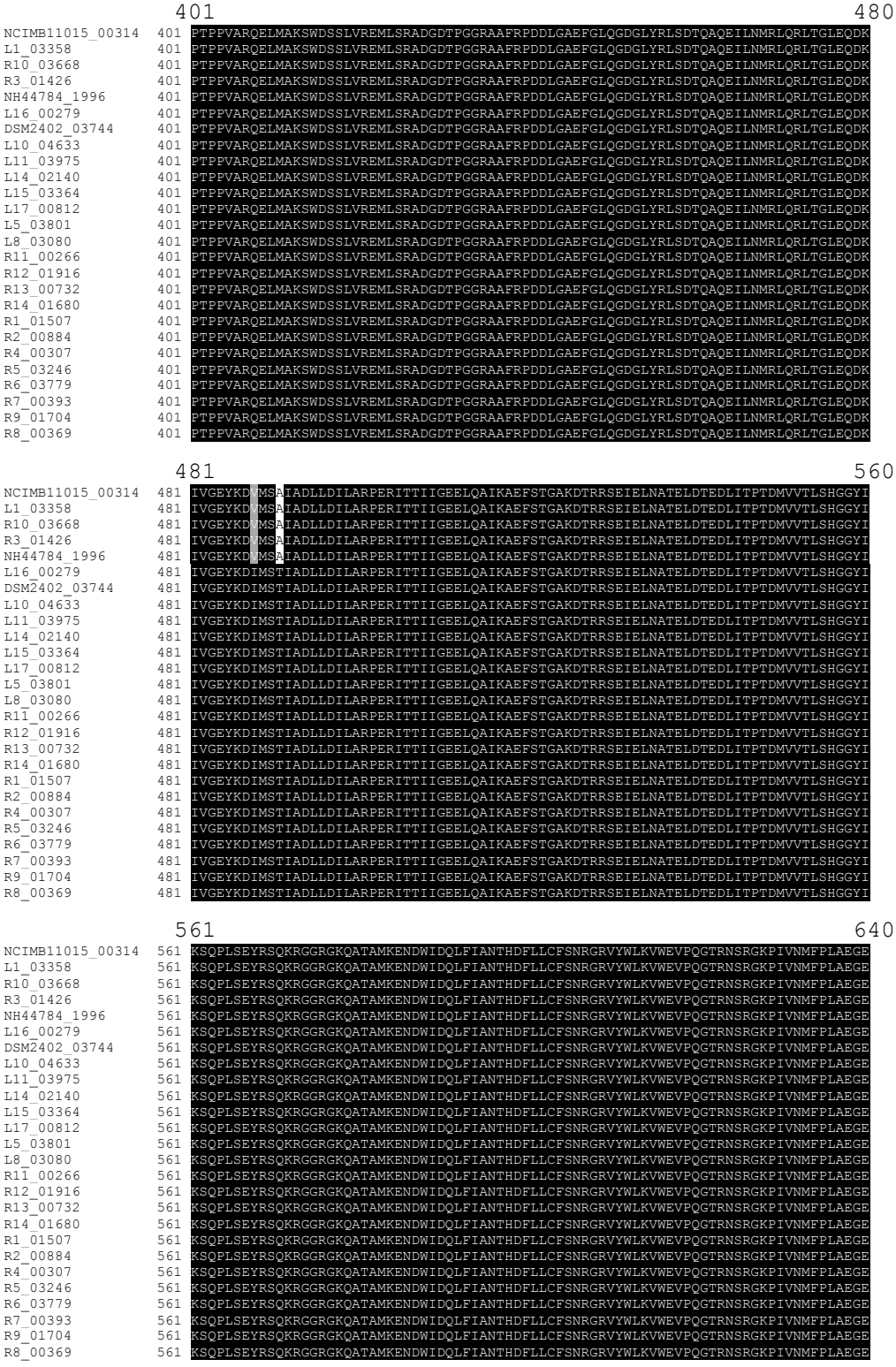
```
RA-41520608  1041 CCGAGATCGCCCAGTTTTTCGCCGGAGCCCTGGCCGAGCCCAACGAAGTGCTGGTAGCCCATGATGATGCGGGGGCCGTT

P.aeruginosa 1121 GTTGGGCATGTCGAGTTATCCATCCGCGAGGATGTCGCAGGGCTGGAAGGCATCAGAGCGGGCTATATCGAAGGCCTGTA
R4           1121 GTTGGGCATGTCGAGTTATCCATCCGCGAGGATGTCGCAGGGCTGGAAGGCATCAGAGCGGGCTATATCGAAGGCCTGTA
R8           1121 GTTGGGCATGTCGAGTTATCCATCCGCGAGGATGTCGCAGGGCTGGAAGGCATCAGAGCGGGCTATATCGAAGGCCTGTA
Aci_16       1121 GTTGGGCATGTCGAGTTATCCATCCGCGAGGATGTCGCAGGGCTGGAAGGCATCAGAGCGGGCTATATCGAAGGCCTGTA
RA-41520608  1121 GTTGGGCATGTCGAGTTATCCATCCGCGAGGATGTCGCAGGGCTGGAAGGCATCAGAGCGGGCTATATCGAAGGCCTGTA

P.aeruginosa 1201 CATCGAGGAGGCCCATCGCTCGTCCAGCGTCGCGACGCAGTTACTACGGCACTCCGAGCAATGGGCCCAAAGTCAGGGAT
R4           1201 CATCGAGGAGGCCCATCGCTCGTCCAGCGTCGCGACGCAGTTACTACGGCACTCCGAGCAATGGGCCCAAAGTCAGGGAT
R8           1201 CATCGAGGAGGCCCATCGCTCGTCCAGCGTCGCGACGCAGTTACTACGGCACTCCGAGCAATGGGCCCAAAGTCAGGGAT
Aci_16       1201 CATCGAGGAGGCCCATCGCTCGTCCAGCGTCGCGACGCAGTTACTACGGCACTCCGAGCAATGGGCCCAAAGTCAGGGAT
RA-41520608  1201 CATCGAGGAGGCCCATCGCTCGTCCAGCGTCGCGACGCAGTTACTACGGCACTCCGAGCAATGGGCCCAAAGTCAGGGAT

P.aeruginosa 1281 GCCGGGCGTTTGCATCGGATCGAGAGGATCGCCTGATCATCCATAAGCGGTTTTCTGTGAGTCCGCTTTCTAACCCTTCA
R4           1281 GCCGGGCGTTTGCATCGGATCGAGAGGATCGCCTGATCATCCATAAGCGGTTTTCTGTGAGTCCGCTTTCTAACCCTTCA
R8           1281 GCCGGGCGTTTGCATCGGATCGAGAGGATCGCCTGATCATCCATAAGCGGTTTTCTGTGAGTCCGCTTTCTAACCCTTCA
Aci_16       1281 GCCGGGCGTTTGCATCGGATCGAGAGGATCGCCTGATCATCCATAAGCGGTTTTCTGTGAGTCCGCTTTCTAACCCTTCA
RA-41520608  1281 GCCGGGCGTTTGCATCGGATCGAGAGGATCGCCTGATCATCCATAAGCGGTTTTCTGTGAGTCCGCTTTCTAACCCTTCA

P.aeruginosa 1361 TTCCAGCGGACCGCCTTCGGCGGCCGCTGAATTT
R4           1361 TTCCAGCGGACCGCCTTCGGCGGCCGCTGAATTT
R8           1361 TTCCAGCGGACCGCCTTCGGCGGCCGCTGAATTT
Aci_16       1361 TTCCAGCGGACCGCCTTCGGCGGCCGCTGAATTT
RA-41520608  1361 TTCCAGCGGACCGCCTTCGGCGGCCGCTGAAT
```

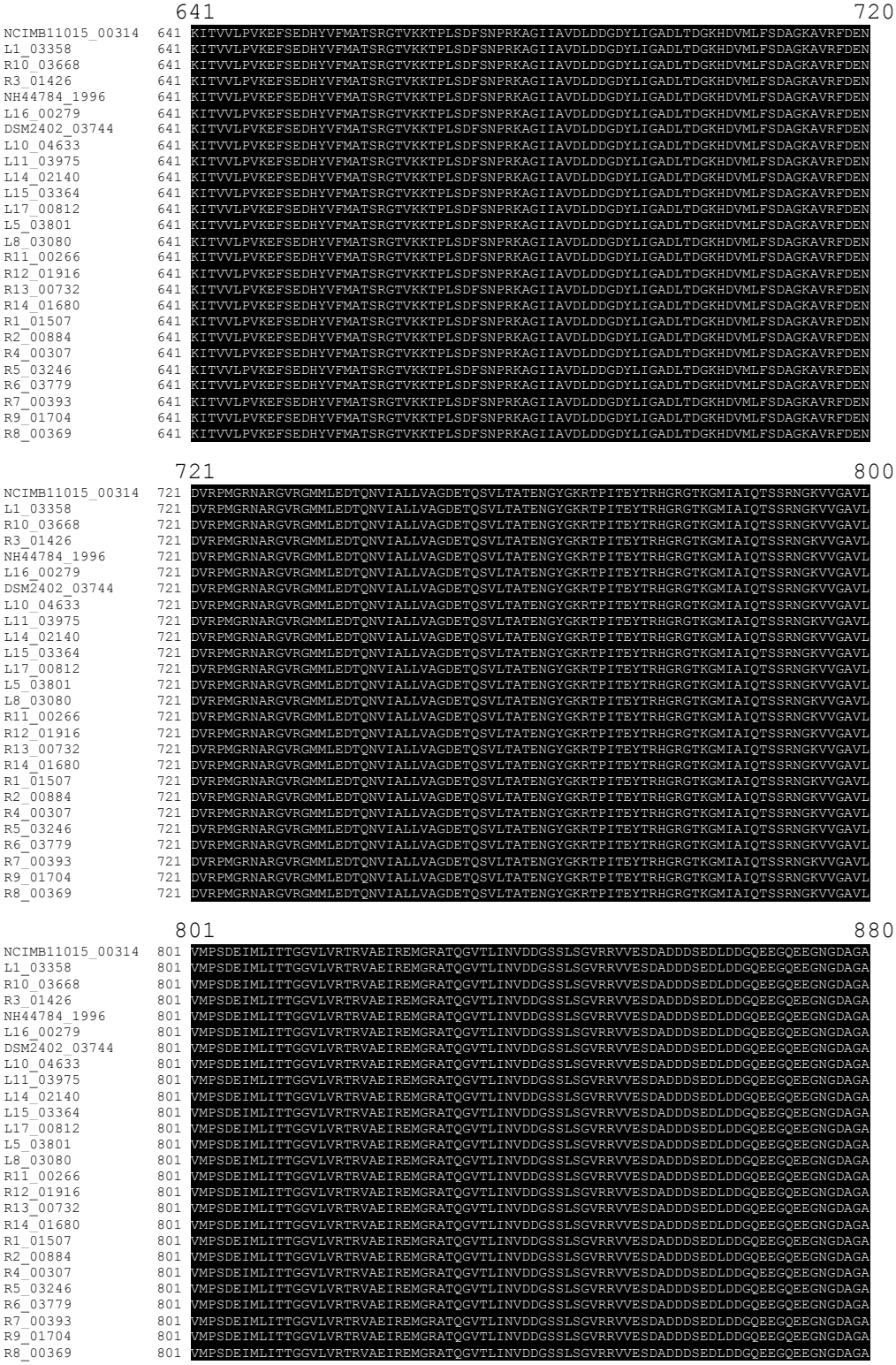
Figure A3.6: Multiple sequence alignment to show amino acid sequence alignment of topoisomerase extracted from *A. xylosoxidans*; (A) DNA Gyrase subunit A, (B) DNA Gyrase subunit B, (C) Topoisomerase IV subunit A and (D) Topoisomerase IV subunit B. The substitution of amino acid at QRDR is marked by an asterisk.

A) DNA Gyrase subunit A




161 240
NCIMB11015_00314 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L1_03358 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R10_03668 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R3_01426 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
NH44784_1996 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L16_00279 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
DSM2402_03744 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L10_04633 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L11_03975 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L14_02140 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L15_03364 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L17_00812 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L5_03801 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
L8_03080 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R11_00266 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R12_01916 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R13_00732 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R14_01680 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R1_01507 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R2_00884 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R4_00307 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R5_03246 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R6_03779 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R7_00393 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R9_01704 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
R8_00369 161 SRLPNLLVNGSSGIAVGMATNIPPHNLQEVVDGCLYCLRNPACSVDELMEIIPAPDFPTGGVIYGMSGVREGYRTGRGRV
241 320
NCIMB11015_00314 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L1_03358 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R10_03668 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R3_01426 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
NH44784_1996 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L16_00279 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
DSM2402_03744 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L10_04633 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L11_03975 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L14_02140 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L15_03364 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L17_00812 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L5_03801 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
L8_03080 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R11_00266 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R12_01916 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R13_00732 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R14_01680 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R1_01507 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R2_00884 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R4_00307 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R5_03246 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R6_03779 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R7_00393 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R9_01704 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
R8_00369 241 IMRAKTHFEDMEKGNRQAIVVDAIPYQVNKKTLQERIAELVNDKKIEGISDIRDESDKDGMRLVIELKRGEVPEVVLNNL
321 400
NCIMB11015_00314 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L1_03358 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R10_03668 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R3_01426 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
NH44784_1996 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L16_00279 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
DSM2402_03744 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L10_04633 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L11_03975 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L14_02140 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L15_03364 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L17_00812 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L5_03801 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
L8_03080 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R11_00266 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R12_01916 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R13_00732 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R14_01680 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R1_01507 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R2_00884 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R4_00307 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R5_03246 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R6_03779 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R7_00393 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R9_01704 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA
R8_00369 321 YKNTQLQDTFGMNLVALVDGQPRLLNLKQMIDYFLQHRREVVTRRTVFQLRKARERGHVLEGLAVALANIDDFIAIIKAA

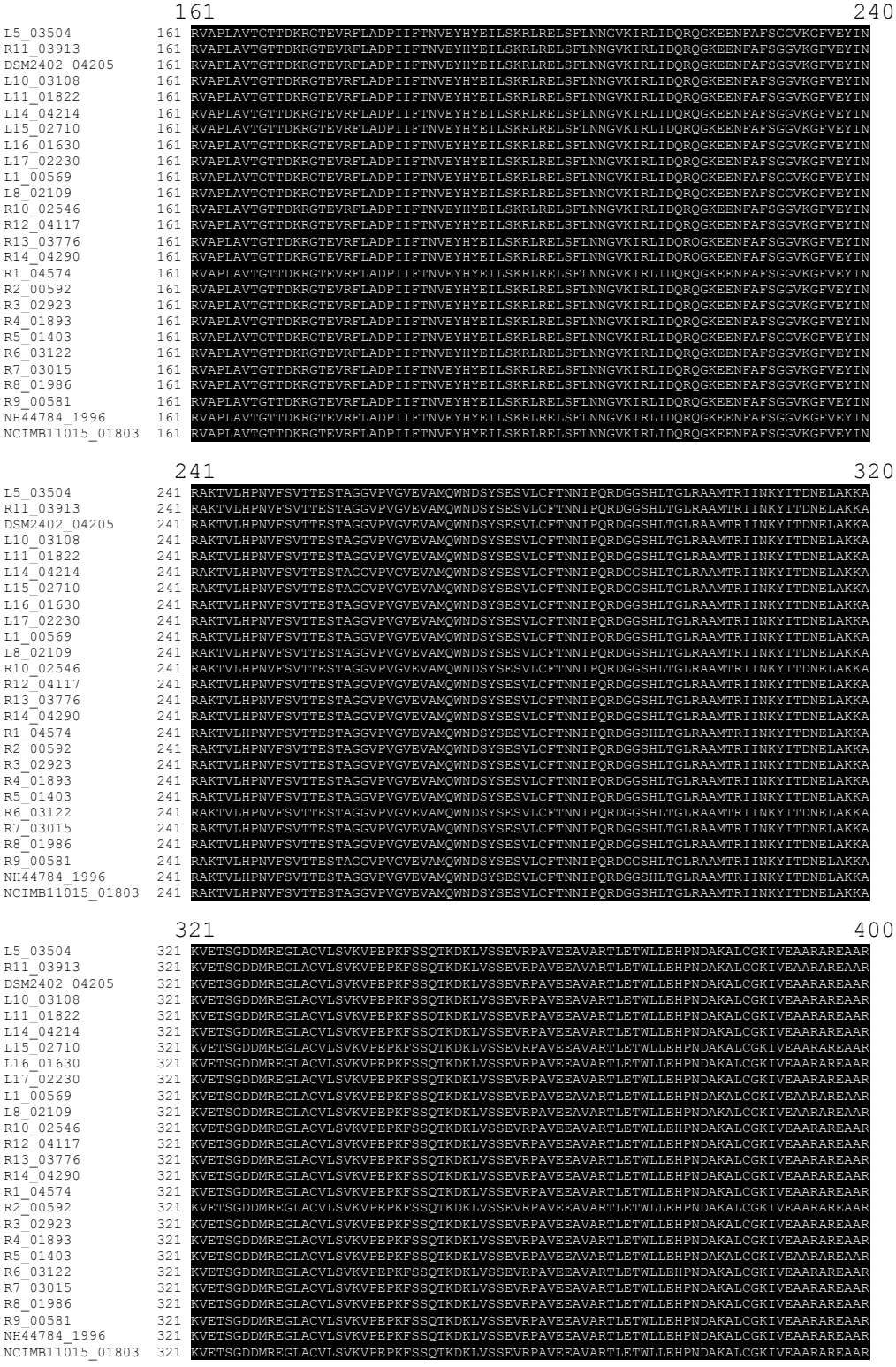

401 480
NCIMB11015_00314 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L1_03358 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R10_03668 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R3_01426 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
NH44784_1996 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L16_00279 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
DSM2402_03744 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L10_04633 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L11_03975 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L14_02140 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L15_03364 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L17_00812 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L5_03801 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
L8_03080 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R11_00266 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R12_01916 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R13_00732 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R14_01680 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R1_01507 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R2_00884 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R4_00307 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R5_03246 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R6_03779 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R7_00393 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R9_01704 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
R8_00369 401 PTPPVARQELMAKSWDSSLVREMLSRADGDTPGGRAAFRPDDLGAEFGLQGDGLYRLSDTQAQEILNMRLQRLTGLEQDK
481 560
NCIMB11015_00314 481 IVGEYKDIMSAIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L1_03358 481 IVGEYKDIMSAIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R10_03668 481 IVGEYKDIMSAIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R3_01426 481 IVGEYKDIMSAIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
NH44784_1996 481 IVGEYKDIMSAIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L16_00279 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
DSM2402_03744 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L10_04633 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L11_03975 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L14_02140 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L15_03364 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L17_00812 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L5_03801 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
L8_03080 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R11_00266 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R12_01916 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R13_00732 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R14_01680 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R1_01507 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R2_00884 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R4_00307 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R5_03246 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R6_03779 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R7_00393 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R9_01704 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
R8_00369 481 IVGEYKDIMSTIADLLDILARPERITTIIGEELQAIKAEFSTGAKDTRRSEIELNATELDTEDLITPTDMVVTLSHGGYI
561 640
NCIMB11015_00314 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L1_03358 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R10_03668 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R3_01426 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
NH44784_1996 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L16_00279 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
DSM2402_03744 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L10_04633 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L11_03975 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L14_02140 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L15_03364 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L17_00812 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L5_03801 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
L8_03080 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R11_00266 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R12_01916 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R13_00732 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R14_01680 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R1_01507 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R2_00884 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R4_00307 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R5_03246 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R6_03779 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R7_00393 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R9_01704 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE
R8_00369 561 KSQPLSEYRSQKRGGRGKQATAMKENDWIDQLFIANTHDFLLCFSNRGRVYWLKVWEVPQGTRNSRGKPIVNMFPLAEGE

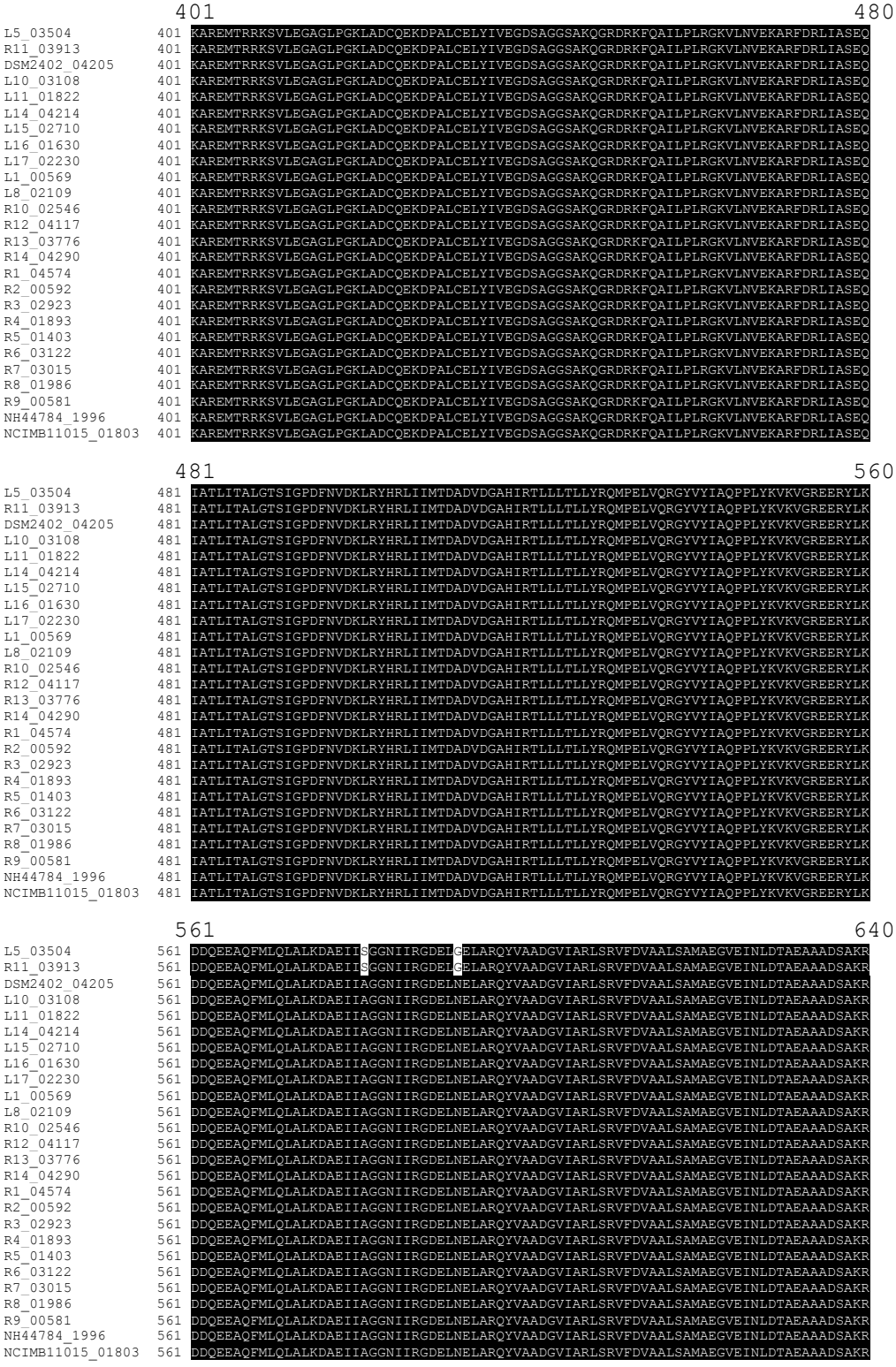

641 720
NCIMB11015_00314 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L1_03358 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R10_03668 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R3_01426 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
NH44784_1996 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L16_00279 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
DSM2402_03744 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L10_04633 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L11_03975 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L14_02140 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L15_03364 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L17_00812 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L5_03801 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
L8_03080 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R11_00266 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R12_01916 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R13_00732 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R14_01680 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R1_01507 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R2_00884 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R4_00307 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R5_03246 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R6_03779 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R7_00393 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R9_01704 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
R8_00369 641 KITVVLPVKEFSEDHYVFMATSRGTVKKTPLSDFSNPRKAGIIAVDLDDGDYLIGADLTDGKHDVMLFSDAGKAVRFDEN
721 800
NCIMB11015_00314 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L1_03358 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R10_03668 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R3_01426 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
NH44784_1996 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L16_00279 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
DSM2402_03744 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L10_04633 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L11_03975 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L14_02140 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L15_03364 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L17_00812 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L5_03801 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
L8_03080 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R11_00266 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R12_01916 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R13_00732 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R14_01680 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R1_01507 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R2_00884 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R4_00307 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R5_03246 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R6_03779 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R7_00393 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R9_01704 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
R8_00369 721 DVRPMGRNARGVRGMMLEDTQNVIALLVAGDETQSVLTATENGYGKRTPITEYTRHGRGTKGMIAIQTSSRNGKVVGAVL
801 880
NCIMB11015_00314 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L1_03358 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R10_03668 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R3_01426 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
NH44784_1996 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L16_00279 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
DSM2402_03744 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L10_04633 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L11_03975 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L14_02140 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L15_03364 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L17_00812 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L5_03801 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
L8_03080 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R11_00266 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R12_01916 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R13_00732 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R14_01680 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R1_01507 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R2_00884 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R4_00307 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R5_03246 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R6_03779 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R7_00393 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R9_01704 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA
R8_00369 801 VMPSDEIMLITTGGVLVRTRVAEIREMGRATQGVTLINVDDGSSLSGVRRVVESDADDDSEDLDDGQEEGQEEGNGDAGA

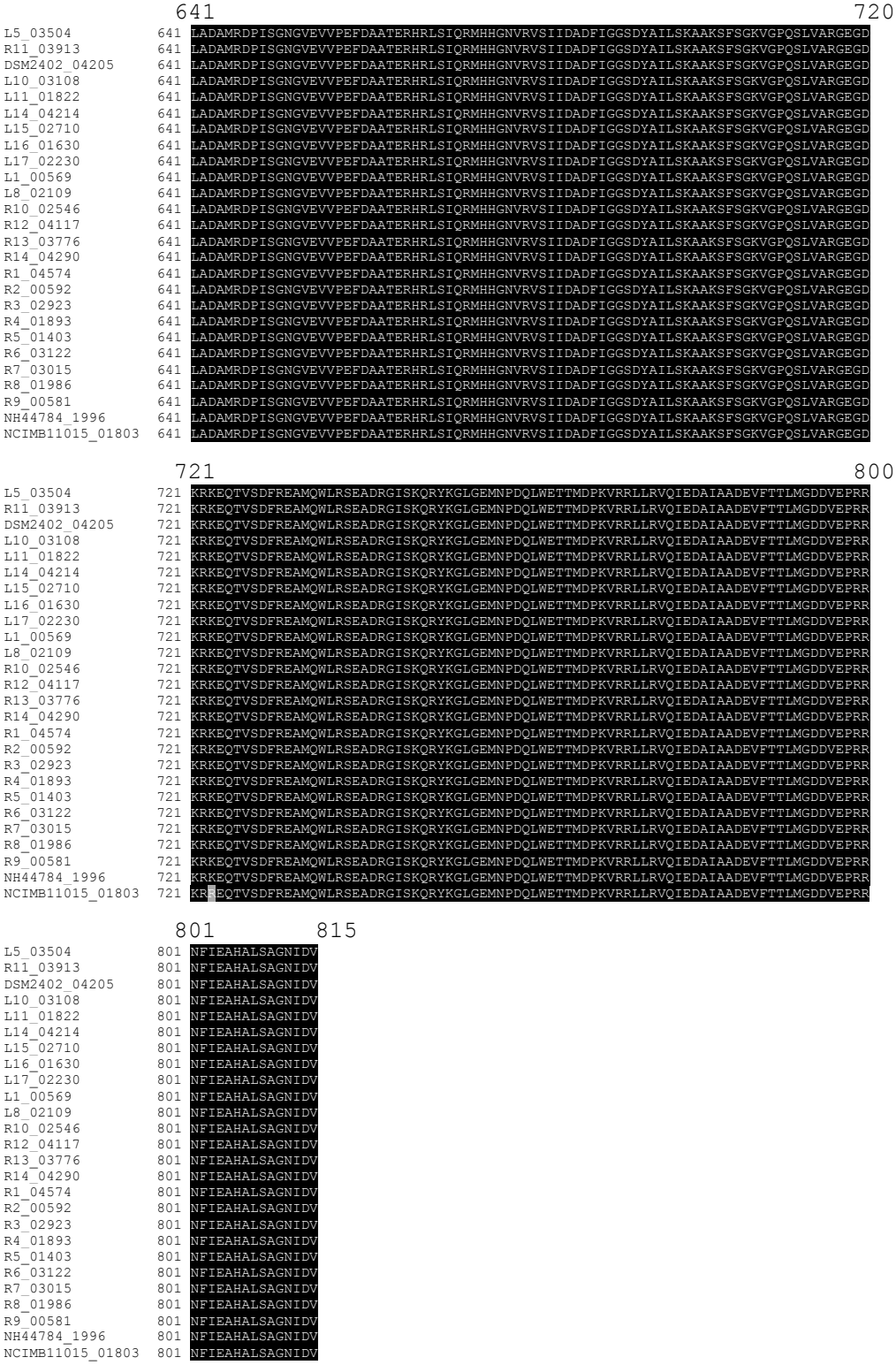
```
                    881    889
NCIMB11015_00314    881 TDSTEPTEQ
L1_03358            881 TDSTEPTEQ
R10_03668           881 TDSTEPTEQ
R3_01426            881 TDSTEPTEQ
NH44784_1996        881 TDSTEPTEQ
L16_00279           881 MDSTEPTEQ
DSM2402_03744       881 TDSTEPTEQ
L10_04633           881 TDSTEPTEQ
L11_03975           881 TDSTEPTEQ
L14_02140           881 TDSTEPTEQ
L15_03364           881 TDSTEPTEQ
L17_00812           881 TDSTEPTEQ
L5_03801            881 TDSTEPTEQ
L8_03080            881 TDSTEPTEQ
R11_00266           881 TDSTEPTEQ
R12_01916           881 TDSTEPTEQ
R13_00732           881 TDSTEPTEQ
R14_01680           881 TDSTEPTEQ
R1_01507            881 TDSTEPTEQ
R2_00884            881 TDSTEPTEQ
R4_00307            881 TDSTEPTEQ
R5_03246            881 TDSTEPTEQ
R6_03779            881 TDSTEPTEQ
R7_00393            881 TDSTEPTEQ
R9_01704            881 TDSTEPTEQ
R8_00369            881 TDSTEPTEQ
```

B) DNA Gyrase subunit B

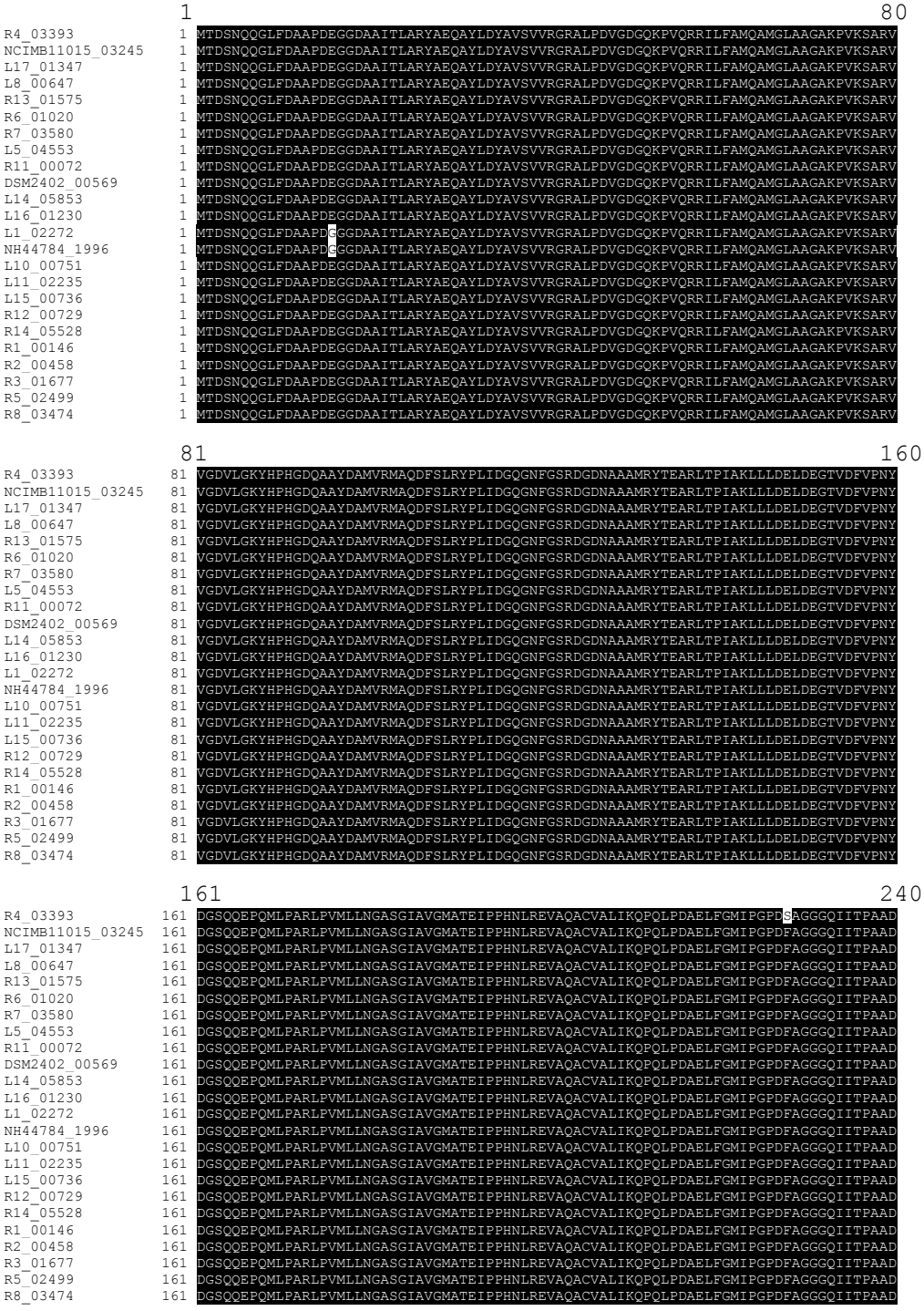
```
                      1                                                                                80
L5_03504              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R11_03913             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
DSM2402_04205         1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L10_03108             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L11_01822             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L14_04214             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L15_02710             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L16_01630             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L17_02230             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L1_00569              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
L8_02109              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R10_02546             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R12_04117             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R13_03776             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R14_04290             1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R1_04574              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R2_00592              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R3_02923              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R4_01893              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R5_01403              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R6_03122              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R7_03015              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R8_01986              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
R9_00581              1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
NH44784_1996          1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV
NCIMB11015_01803      1 MSDQQNTTPENGGYGADSIKMLKGLEAVRKRPGMYIGDTSDGTGLHHMVFEVVDNAIDEALAGYCDDIVVTIHTDNSISV

                     81                                                                               160
L5_03504             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R11_03913            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
DSM2402_04205        81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L10_03108            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L11_01822            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L14_04214            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L15_02710            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L16_01630            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L17_02230            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L1_00569             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
L8_02109             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R10_02546            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R12_04117            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R13_03776            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R14_04290            81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R1_04574             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R2_00592             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R3_02923             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R4_01893             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R5_01403             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R6_03122             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R7_03015             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R8_01986             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
R9_00581             81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
NH44784_1996         81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
NCIMB11015_01803     81 TDNGRGIPTDIHKDDEFHRSAAEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSEWLRLTIRRNGEVHQMEFRQGE
```

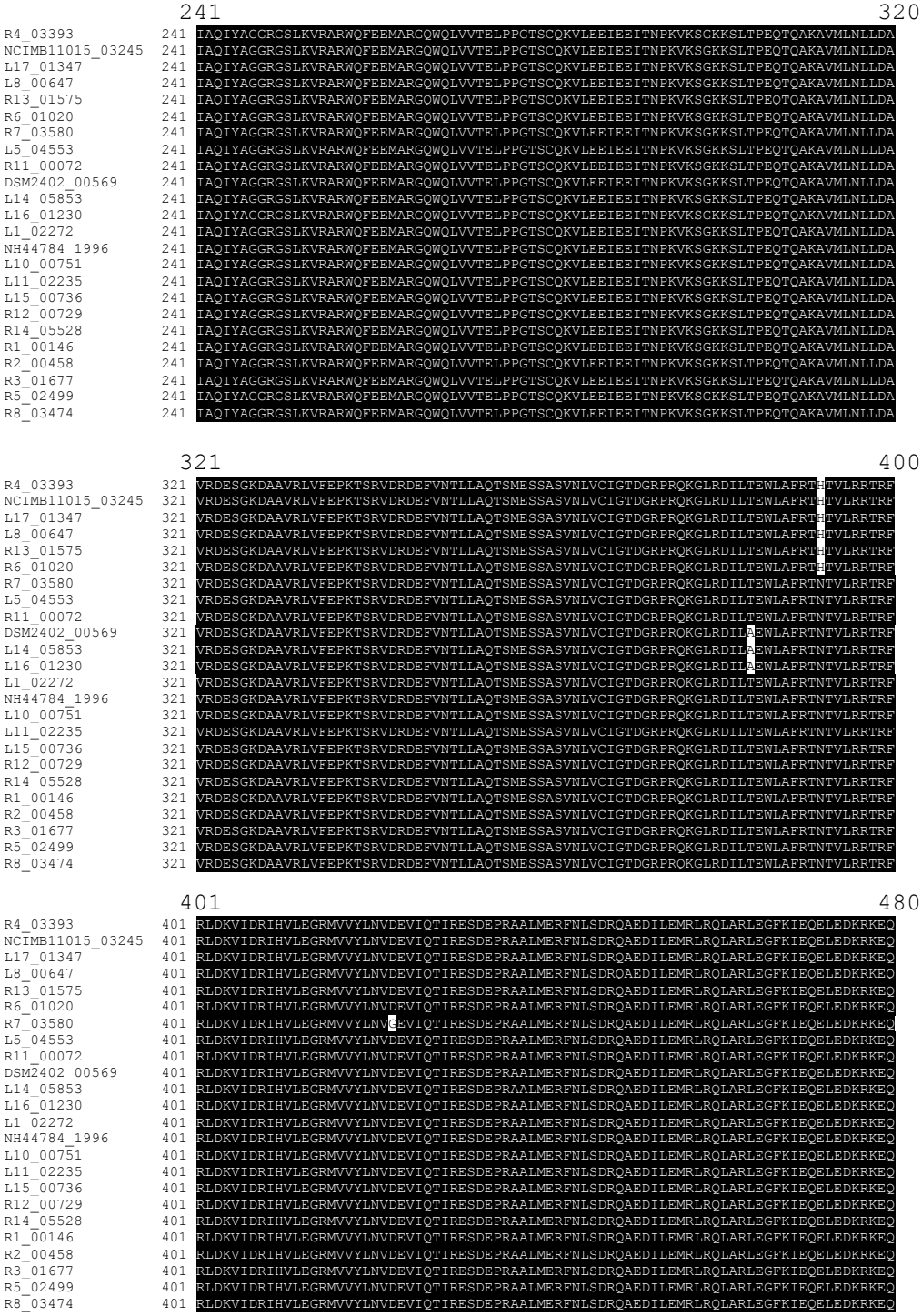

161 240
L5_03504 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R11_03913 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
DSM2402_04205 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L10_03108 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L11_01822 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L14_04214 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L15_02710 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L16_01630 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L17_02230 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L1_00569 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
L8_02109 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R10_02546 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R12_04117 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R13_03776 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R14_04290 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R1_04574 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R2_00592 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R3_02923 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R4_01893 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R5_01403 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R6_03122 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R7_03015 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R8_01986 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
R9_00581 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
NH44784_1996 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
NCIMB11015_01803 161 RVAPLAVTGTTDKRGTEVRFLADPIIFTNVEYHYEILSKRLRELSFLNNGVKIRLIDQRQGKEENFAFSGGVKGFVEYIN
241 320
L5_03504 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R11_03913 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
DSM2402_04205 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L10_03108 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L11_01822 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L14_04214 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L15_02710 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L16_01630 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L17_02230 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L1_00569 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
L8_02109 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R10_02546 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R12_04117 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R13_03776 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R14_04290 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R1_04574 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R2_00592 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R3_02923 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R4_01893 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R5_01403 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R6_03122 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R7_03015 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R8_01986 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
R9_00581 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
NH44784_1996 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
NCIMB11015_01803 241 RAKTVLHPNVFSVTTESTAGGVPVGVEVAMQWNDSYSESVLCFTNNIPQRDGGSHLTGLRAAMTRIINKYITDNELAKKA
321 400
L5_03504 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R11_03913 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
DSM2402_04205 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L10_03108 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L11_01822 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L14_04214 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L15_02710 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L16_01630 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L17_02230 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L1_00569 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
L8_02109 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R10_02546 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R12_04117 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R13_03776 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R14_04290 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R1_04574 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R2_00592 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R3_02923 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R4_01893 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R5_01403 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R6_03122 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R7_03015 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R8_01986 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
R9_00581 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
NH44784_1996 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR
NCIMB11015_01803 321 KVETSGDDMREGLACVLSVKVPEPKFSSQTKDKLVSSEVRPAVEEAVARTLETWLLEHPNDAKALCGKIVEAARAREAAR

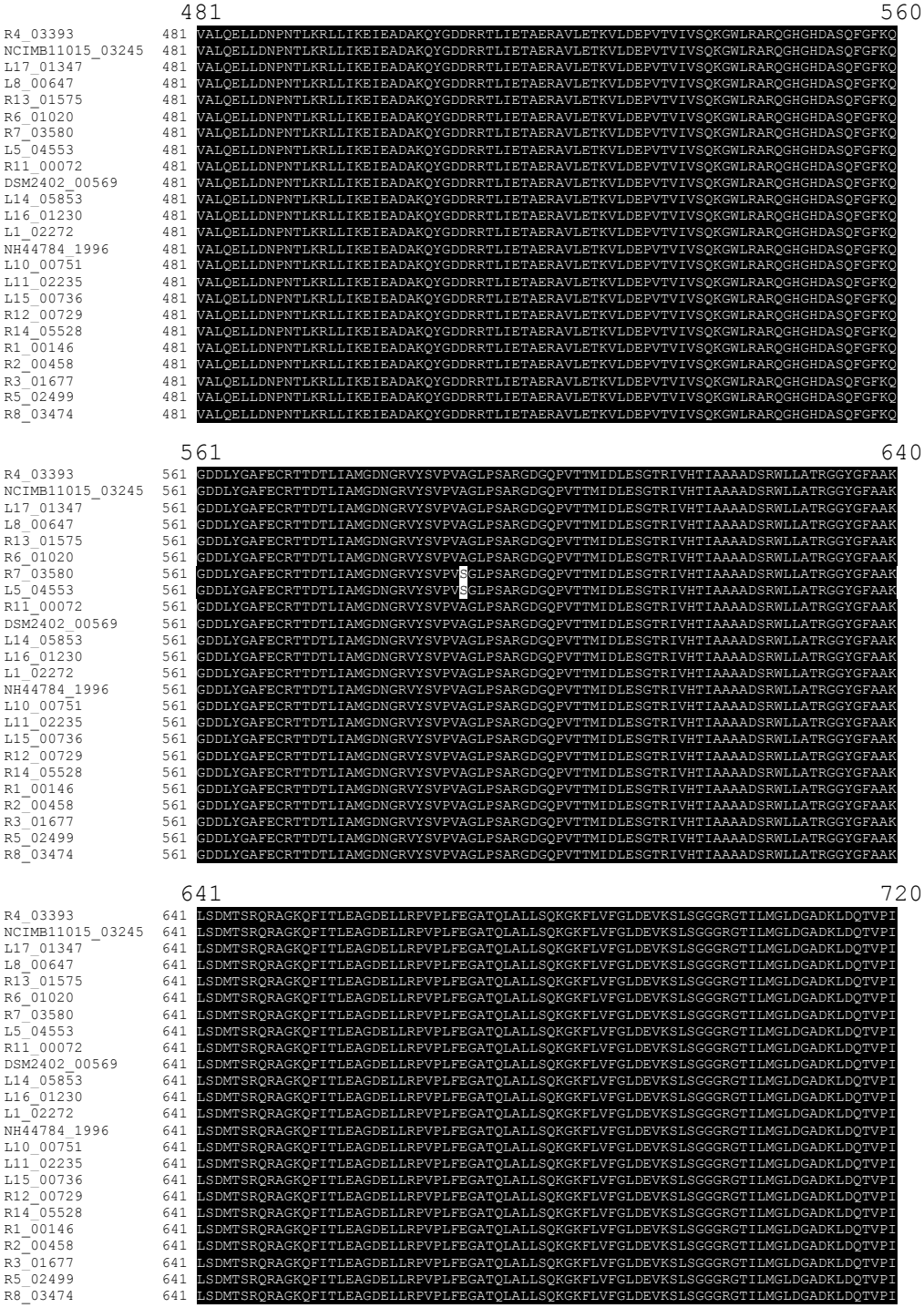

401 480
L5_03504 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R11_03913 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
DSM2402_04205 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L10_03108 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L11_01822 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L14_04214 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L15_02710 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L16_01630 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L17_02230 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L1_00569 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
L8_02109 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R10_02546 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R12_04117 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R13_03776 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R14_04290 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R1_04574 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R2_00592 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R3_02923 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R4_01893 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R5_01403 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R6_03122 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R7_03015 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R8_01986 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
R9_00581 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
NH44784_1996 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
NCIMB11015_01803 401 KAREMTRRKSVLEGAGLPGKLADCQEKDPALCELYIVEGDSAGGSAKQGRDRKFQAILPLRGKVLNVEKARFDRLIASEQ
481 560
L5_03504 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R11_03913 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
DSM2402_04205 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L10_03108 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L11_01822 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L14_04214 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L15_02710 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L16_01630 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L17_02230 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L1_00569 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
L8_02109 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R10_02546 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R12_04117 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R13_03776 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R14_04290 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R1_04574 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R2_00592 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R3_02923 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R4_01893 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R5_01403 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R6_03122 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R7_03015 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R8_01986 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
R9_00581 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
NH44784_1996 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
NCIMB11015_01803 481 IATLITALGTSIGPDFNVDKLRYHRLIIMTDADVDGAHIRTLLLTLLYRQMPELVQRGYVYIAQPPLYKVKVGREERYLK
561 640
L5_03504 561 DDQEEAQFMLQLALKDAEIISGGNIIRGDELGELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R11_03913 561 DDQEEAQFMLQLALKDAEIISGGNIIRGDELGELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
DSM2402_04205 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L10_03108 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L11_01822 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L14_04214 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L15_02710 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L16_01630 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L17_02230 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L1_00569 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
L8_02109 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R10_02546 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R12_04117 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R13_03776 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R14_04290 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R1_04574 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R2_00592 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R3_02923 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R4_01893 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R5_01403 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R6_03122 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R7_03015 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R8_01986 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
R9_00581 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
NH44784_1996 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR
NCIMB11015_01803 561 DDQEEAQFMLQLALKDAEIIAGGNIIRGDELNELARQYVAADGVIARLSRVFDVAALSAMAEGVEINLDTAEAAADSAKR

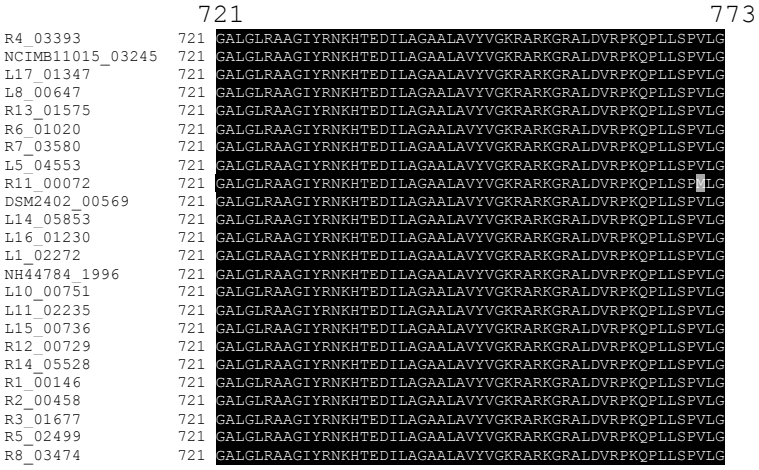

641 720
L5_03504 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R11_03913 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
DSM2402_04205 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L10_03108 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L11_01822 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L14_04214 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L15_02710 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L16_01630 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L17_02230 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L1_00569 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
L8_02109 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R10_02546 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R12_04117 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R13_03776 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R14_04290 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R1_04574 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R2_00592 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R3_02923 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R4_01893 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R5_01403 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R6_03122 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R7_03015 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R8_01986 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
R9_00581 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
NH44784_1996 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
NCIMB11015_01803 641 LADAMRDPISGNGVEVVPEFDAATERHRLSIQRMHHGNVRVSIIDADFIGGSDYAILSKAAKSFSGKVGPQSLVARGEGD
721 800
L5_03504 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R11_03913 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
DSM2402_04205 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L10_03108 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L11_01822 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L14_04214 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L15_02710 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L16_01630 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L17_02230 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L1_00569 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
L8_02109 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R10_02546 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R12_04117 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R13_03776 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R14_04290 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R1_04574 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R2_00592 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R3_02923 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R4_01893 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R5_01403 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R6_03122 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R7_03015 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R8_01986 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
R9_00581 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
NH44784_1996 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
NCIMB11015_01803 721 KRKEQTVSDFREAMQWLRSEADRGISKQRYKGLGEMNPDQLWETTMDPKVRRLLRVQIEDAIAADEVFTTLMGDDVEPRR
801 815
L5_03504 801 NFIEAHALSAGNIDV
R11_03913 801 NFIEAHALSAGNIDV
DSM2402_04205 801 NFIEAHALSAGNIDV
L10_03108 801 NFIEAHALSAGNIDV
L11_01822 801 NFIEAHALSAGNIDV
L14_04214 801 NFIEAHALSAGNIDV
L15_02710 801 NFIEAHALSAGNIDV
L16_01630 801 NFIEAHALSAGNIDV
L17_02230 801 NFIEAHALSAGNIDV
L1_00569 801 NFIEAHALSAGNIDV
L8_02109 801 NFIEAHALSAGNIDV
R10_02546 801 NFIEAHALSAGNIDV
R12_04117 801 NFIEAHALSAGNIDV
R13_03776 801 NFIEAHALSAGNIDV
R14_04290 801 NFIEAHALSAGNIDV
R1_04574 801 NFIEAHALSAGNIDV
R2_00592 801 NFIEAHALSAGNIDV
R3_02923 801 NFIEAHALSAGNIDV
R4_01893 801 NFIEAHALSAGNIDV
R5_01403 801 NFIEAHALSAGNIDV
R6_03122 801 NFIEAHALSAGNIDV
R7_03015 801 NFIEAHALSAGNIDV
R8_01986 801 NFIEAHALSAGNIDV
R9_00581 801 NFIEAHALSAGNIDV
NH44784_1996 801 NFIEAHALSAGNIDV
NCIMB11015_01803 801 NFIEAHALSAGNIDV

C) Topoisomerase IV subunit A

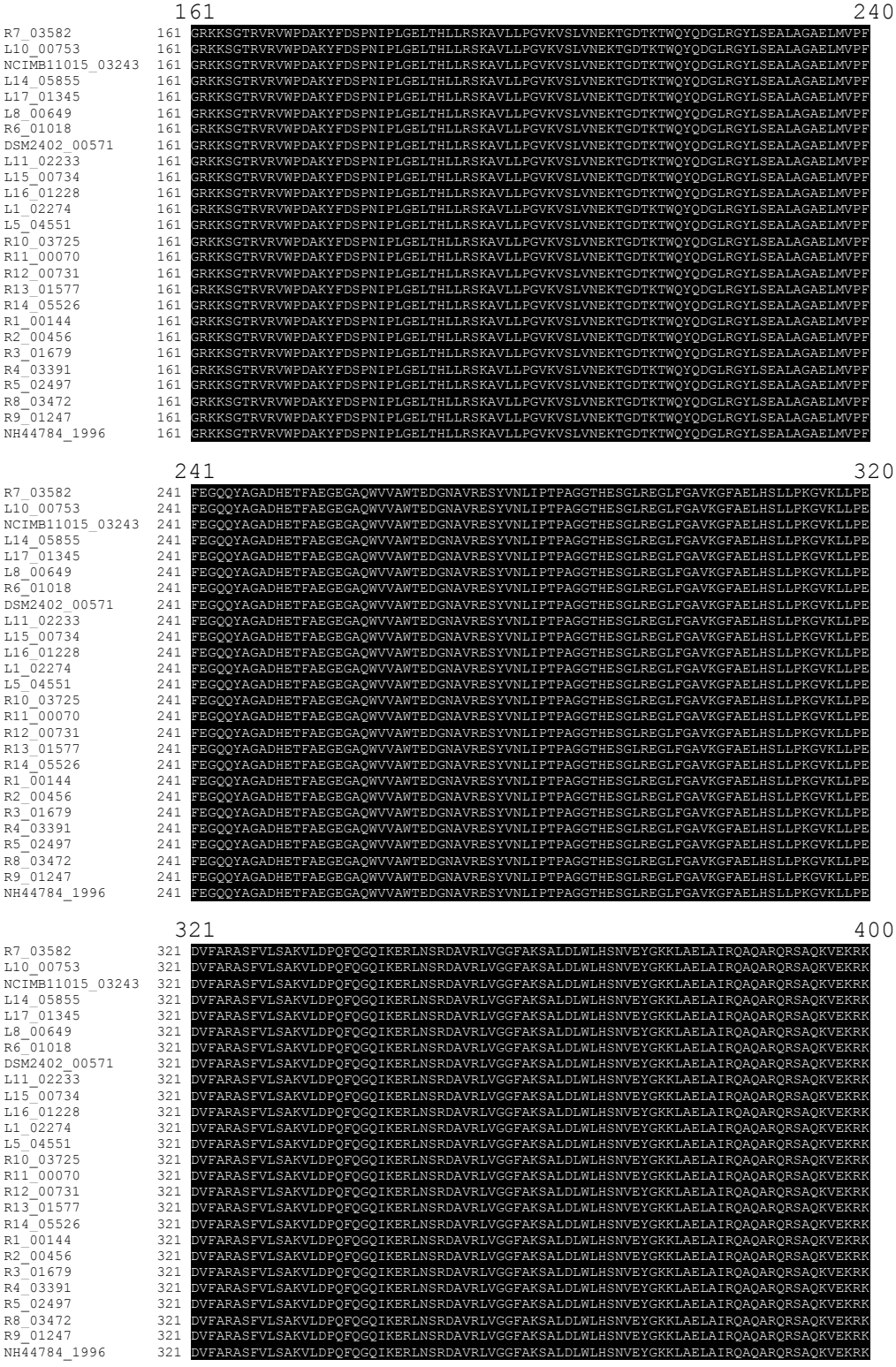


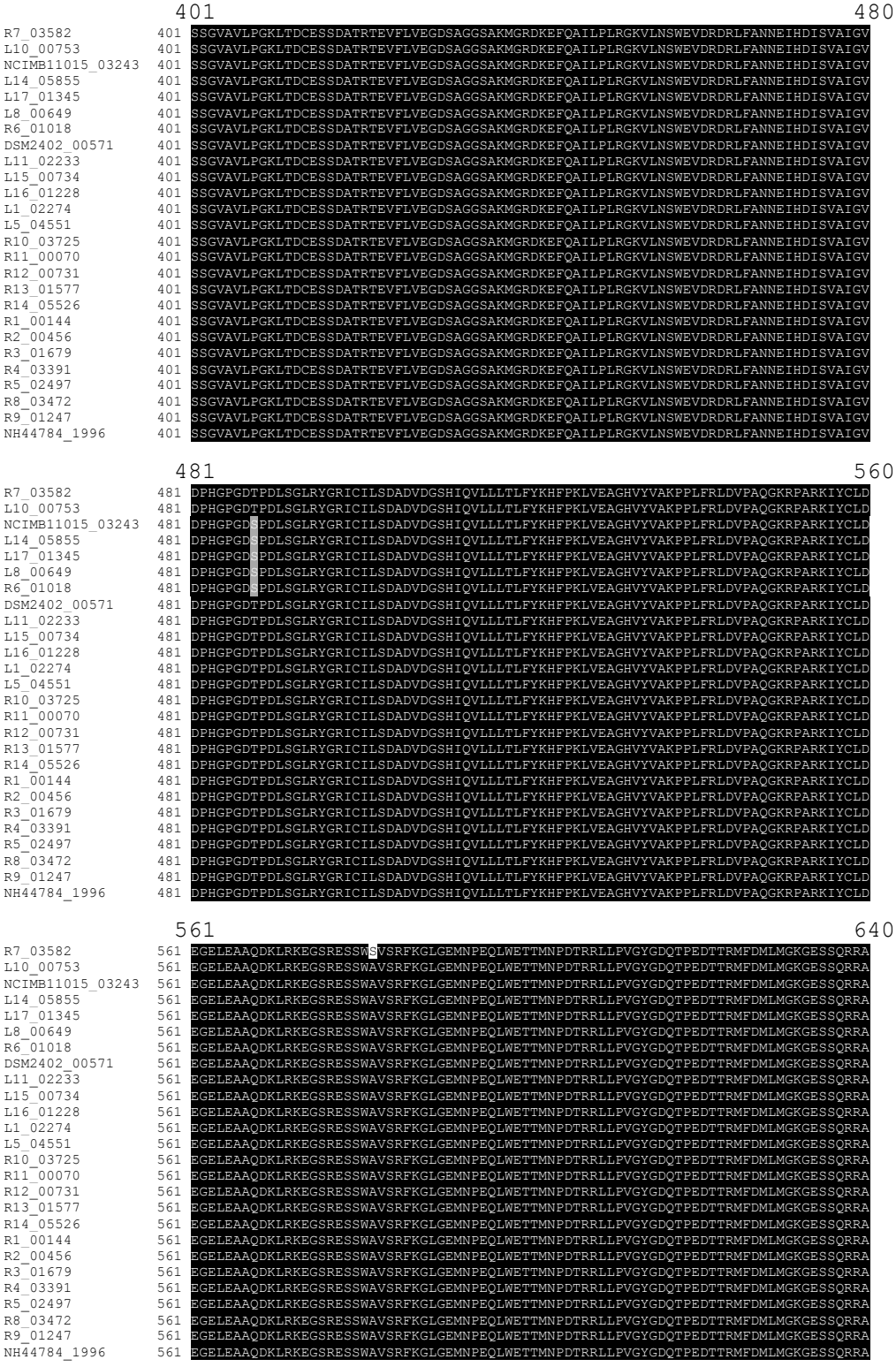

241 320
R4_03393 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
NCIMB11015_03245 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L17_01347 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L8_00647 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R13_01575 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R6_01020 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R7_03580 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L5_04553 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R11_00072 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
DSM2402_00569 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L14_05853 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L16_01230 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L1_02272 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
NH44784_1996 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L10_00751 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L11_02235 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
L15_00736 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R12_00729 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R14_05528 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R1_00146 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R2_00458 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R3_01677 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R5_02499 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
R8_03474 241 IAQIYAGGRGSLKVRARWQFEEMARGQWQLVVTELPPGTSCQKVLEEIEEITNPKVKSGKKSLTPEQTQAKAVMLNLLDA
321 400
R4_03393 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTHTVLRRTRF
NCIMB11015_03245 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTHTVLRRTRF
L17_01347 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTHTVLRRTRF
L8_00647 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTHTVLRRTRF
R13_01575 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTHTVLRRTRF
R6_01020 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTHTVLRRTRF
R7_03580 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
L5_04553 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R11_00072 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
DSM2402_00569 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILAEWLAFRTNTVLRRTRF
L14_05853 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILAEWLAFRTNTVLRRTRF
L16_01230 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILAEWLAFRTNTVLRRTRF
L1_02272 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
NH44784_1996 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
L10_00751 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
L11_02235 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
L15_00736 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R12_00729 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R14_05528 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R1_00146 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R2_00458 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R3_01677 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R5_02499 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
R8_03474 321 VRDESGKDAAVRLVFEPKTSRVDRDEFVNTLLAQTSMESSASVNLVCIGTDGRPRQKGLRDILTEWLAFRTNTVLRRTRF
401 480
R4_03393 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
NCIMB11015_03245 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L17_01347 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L8_00647 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R13_01575 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R6_01020 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R7_03580 401 RLDKVIDRIHVLEGRMVVYLNVEEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L5_04553 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R11_00072 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
DSM2402_00569 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L14_05853 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L16_01230 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L1_02272 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
NH44784_1996 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L10_00751 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L11_02235 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
L15_00736 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R12_00729 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R14_05528 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R1_00146 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R2_00458 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R3_01677 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R5_02499 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ
R8_03474 401 RLDKVIDRIHVLEGRMVVYLNVDEVIQTIRESDEPRAALMERFNLSDRQAEDILEMRLRQLARLEGFKIEQELEDKRKEQ


481 560
R4_03393 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
NCIMB11015_03245 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L17_01347 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L8_00647 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R13_01575 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R6_01020 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R7_03580 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L5_04553 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R11_00072 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
DSM2402_00569 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L14_05853 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L16_01230 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L1_02272 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
NH44784_1996 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L10_00751 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L11_02235 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
L15_00736 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R12_00729 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R14_05528 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R1_00146 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R2_00458 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R3_01677 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R5_02499 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
R8_03474 481 VALQELLDNPNTLKRLLIKEIEADAKQYGDDRRTLIETAERAVLETKVLDEPVTVIVSQKGWLRARQGHGHDASQFGFKQ
561 640
R4_03393 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
NCIMB11015_03245 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L17_01347 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L8_00647 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R13_01575 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R6_01020 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R7_03580 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVSGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L5_04553 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVSGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R11_00072 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
DSM2402_00569 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L14_05853 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L16_01230 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L1_02272 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
NH44784_1996 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L10_00751 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L11_02235 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
L15_00736 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R12_00729 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R14_05528 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R1_00146 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R2_00458 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R3_01677 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R5_02499 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
R8_03474 561 GDDLYGAFECRTTDTLIAMGDNGRVYSVPVAGLPSARGDGQPVTTMIDLESGTRIVHTIAAAADSRWLLATRGGYGFAAK
641 720
R4_03393 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
NCIMB11015_03245 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L17_01347 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L8_00647 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R13_01575 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R6_01020 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R7_03580 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L5_04553 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R11_00072 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
DSM2402_00569 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L14_05853 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L16_01230 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L1_02272 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
NH44784_1996 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L10_00751 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L11_02235 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
L15_00736 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R12_00729 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R14_05528 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R1_00146 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R2_00458 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R3_01677 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R5_02499 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI
R8_03474 641 LSDMTSRQRAGKQFITLEAGDELLRPVPLFEGATQLALLSQKGKFLVFGLDEVKSLSGGGRGTILMGLDGADKLDQTVPI



D) Topoisomerase IV subunit B


161 240
R7_03582 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L10_00753 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
NCIMB11015_03243 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L14_05855 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L17_01345 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L8_00649 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R6_01018 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
DSM2402_00571 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L11_02233 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L15_00734 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L16_01228 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L1_02274 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
L5_04551 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R10_03725 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R11_00070 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R12_00731 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R13_01577 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R14_05526 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R1_00144 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R2_00456 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R3_01679 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R4_03391 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R5_02497 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R8_03472 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
R9_01247 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
NH44784_1996 161 GRKKSGTRVRVWPDAKYFDSPNIPLGELTHLLRSKAVLLPGVKVSLVNEKTGDTKTWQYQDGLRGYLSEALAGAELMVPF
241 320
R7_03582 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L10_00753 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
NCIMB11015_03243 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L14_05855 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L17_01345 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L8_00649 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R6_01018 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
DSM2402_00571 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L11_02233 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L15_00734 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L16_01228 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L1_02274 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
L5_04551 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R10_03725 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R11_00070 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R12_00731 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R13_01577 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R14_05526 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R1_00144 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R2_00456 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R3_01679 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R4_03391 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R5_02497 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R8_03472 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
R9_01247 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
NH44784_1996 241 FEGQQYAGADHETFAEGEGAQWVVAWTEDGNAVRESYVNLIPTPAGGTHESGLREGLFGAVKGFAELHSLLPKGVKLLPE
321 400
R7_03582 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L10_00753 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
NCIMB11015_03243 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L14_05855 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L17_01345 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L8_00649 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R6_01018 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
DSM2402_00571 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L11_02233 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L15_00734 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L16_01228 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L1_02274 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
L5_04551 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R10_03725 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R11_00070 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R12_00731 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R13_01577 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R14_05526 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R1_00144 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R2_00456 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R3_01679 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R4_03391 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R5_02497 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R8_03472 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
R9_01247 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK
NH44784_1996 321 DVFARASFVLSAKVLDPQFQGQIKERLNSRDAVRLVGGFAKSALDLWLHSNVEYGKKLAELAIRQAQARQRSAQKVEKRK


401 480
R7_03582 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L10_00753 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
NCIMB11015_03243 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L14_05855 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L17_01345 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L8_00649 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R6_01018 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
DSM2402_00571 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L11_02233 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L15_00734 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L16_01228 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L1_02274 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
L5_04551 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R10_03725 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R11_00070 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R12_00731 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R13_01577 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R14_05526 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R1_00144 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R2_00456 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R3_01679 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R4_03391 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R5_02497 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R8_03472 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
R9_01247 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
NH44784_1996 401 SSGVAVLPGKLTDCESSDATRTEVFLVEGDSAGGSAKMGRDKEFQAILPLRGKVLNSWEVDRDRLFANNEIHDISVAIGV
481 560
R7_03582 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L10_00753 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
NCIMB11015_03243 481 DPHGPGDSPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L14_05855 481 DPHGPGDSPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L17_01345 481 DPHGPGDSPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L8_00649 481 DPHGPGDSPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R6_01018 481 DPHGPGDSPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
DSM2402_00571 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L11_02233 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L15_00734 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L16_01228 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L1_02274 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
L5_04551 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R10_03725 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R11_00070 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R12_00731 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R13_01577 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R14_05526 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R1_00144 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R2_00456 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R3_01679 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R4_03391 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R5_02497 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R8_03472 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
R9_01247 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
NH44784_1996 481 DPHGPGDTPDLSGLRYGRICILSDADVDGSHIQVLLLTLFYKHFPKLVEAGHVYVAKPPLFRLDVPAQGKRPARKIYCLD
561 640
R7_03582 561 EGELEAAQDKLRKEGSRESSWSVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L10_00753 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
NCIMB11015_03243 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L14_05855 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L17_01345 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L8_00649 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R6_01018 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
DSM2402_00571 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L11_02233 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L15_00734 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L16_01228 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L1_02274 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
L5_04551 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R10_03725 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R11_00070 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R12_00731 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R13_01577 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R14_05526 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R1_00144 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R2_00456 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R3_01679 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R4_03391 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R5_02497 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R8_03472 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
R9_01247 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA
NH44784_1996 561 EGELEAAQDKLRKEGSRESSWAVSRFKGLGEMNPEQLWETTMNPDTRRLLPVGYGDQTPEDTTRMFDMLMGKGESSQRRA

```
                    641       530
R7_03582            641 WIEEKGNLAELDI
L10_00753           641 WIEEKGNLAELDI
NCIMB11015_03243    641 WIEEKGNLAELDI
L14_05855           641 WIEEKGNLAELDI
L17_01345           641 WIEEKGNLAELDI
L8_00649            641 WIEEKGNLAELDI
R6_01018            641 WIEEKGNLAELDI
DSM2402_00571       641 WIEEKGNLAELDI
L11_02233           641 WIEEKGNLAELDI
L15_00734           641 WIEEKGNLAELDI
L16_01228           641 WIEEKGNLAELDI
L1_02274            641 WIEEKGNLAELDI
L5_04551            641 WIEEKGNLAELDI
R10_03725           641 WIEEKGNLAELDI
R11_00070           641 WIEEKGNLAELDI
R12_00731           641 WIEEKGNLAELDI
R13_01577           641 WIEEKGNLAELDI
R14_05526           641 WIEEKGNLAELDI
R1_00144            641 WIEEKGNLAELDI
R2_00456            641 WIEEKGNLAELDI
R3_01679            641 WIEEKGNLAELDI
R4_03391            641 WIEEKGNLAELDI
R5_02497            641 WIEEKGNLAELDI
R8_03472            641 WIEEKGNLAELDI
R9_01247            641 WIEEKGNLAELDI
NH44784_1996        641 WIEEKGNLAELDI
```

Table A3.1: A table to show the zone of inhibition for cephalosporin antibiotic determined on 25 isolates of *A. xylosoxidans*

Isolates	Zone of inhibition (mm)				
	CXM	CTX	CAZ	CRO	FEP
R1	5	11	30	11	31
R2	5	7	30	12	36
R3	5	7	28	11	30
R4	5	5	5	5	19
R5	5	5	27	10	31
R6	5	12	30	12	26
R7	5	12	34	24	11
R8	5	5	21	22	22
R9	5	25	27	24	38
R10	5	9	27	11	31
R11	5	17	34	17	26
R12	5	13	28	12	32
R13	5	5	32	10	30
R14	5	5	30	10	30
L1	5	10	34	15	30
L5	5	7	34	16	25
L8	5	12	32	15	23
L10	5	15	32	20	28
L11	5	19	36	28	30
L14	5	7	33	20	20
L15	5	20	30	24	32
L16	5	14	30	20	17
L17	5	13	30	17	26
NCIMB 11015	5	12	34	16	28
DSM 2402	5	18	32	12	18

Table A3.2: A table to show the zone of inhibition for antibiotic in category: fluoroquinolone and aminoglycoside determined on 25 isolates of *A. xylosoxidans*

Isolates	Zone of inhibition (mm)					
	LEV	CIP	GEN	AMK	SXT	TGC
R1	22	17	9	15	26	30
R2	21	18	11	12	26	28
R3	20	17	10	12	20	24
R4	20	16	5	13	5	25
R5	21	19	5	12	22	29
R6	30	24	12	22	25	32
R7	19	20	5	5	30	20
R8	12	12	16	14	14	22
R9	19	8	8	16	24	26
R10	20	17	10	11	20	24
R11	24	20	15	21	42	30
R12	18	18	9	16	24	25
R13	14	19	5	11	20	25
R14	18	22	9	11	27	26
L1	26	22	15	16	42	30
L5	25	22	12	14	42	27
L8	32	28	18	16	36	39
L10	25	21	13	17	50	30
L11	28	26	20	29	38	25
L14	20	20	5	21	40	25
L15	26	15	14	25	34	26
L16	14	15	4	5	34	24
L17	25	25	25	20	34	33
NCIMB 11015	22	22	14	16	26	32
DSM 2402	12	12	12	11	14	18

Table A3.3: A table to show the zone of inhibition for antibiotic in category: aminopenicillin, ureidopenicillin and carbapenem determined on 25 isolates of *A. xylosoxidans*

Isolates	Zone of inhibition (mm)					
	AMP	AMC	TZP	IPM	MEM	ETP
R1	14	21	38	22	28	32
R2	16	26	38	30	31	36
R3	14	24	36	21	28	32
R4	12	12	38	17	5	12
R5	14	20	39	21	28	30
R6	32	38	54	39	42	50
R7	20	27	50	36	48	32
R8	12	19	22	16	5	13
R9	16	30	38	30	28	38
R10	12	22	32	21	28	32
R11	25	30	50	36	46	42
R12	16	30	34	30	48	36
R13	14	22	32	20	30	30
R14	14	22	38	21	30	30
L1	21	28	48	24	40	40
L5	21	32	53	31	46	44
L8	20	28	46	35	24	26
L10	25	28	48	34	44	42
L11	30	32	50	38	56	48
L14	20	28	48	32	42	44
L15	30	32	52	40	48	46
L16	23	25	44	34	36	30
L17	23	30	44	33	40	43
NCIMB 11015	22	32	38	36	32	32
DSM 2402	12	18	32	34	28	32

Table A3.4: A table to show minimum inhibitory concentration of Thai isolates of *A. xylosoxidans* for antimicrobial agents

Isolates	Minimum inhibitory Concentration (µg/ml)																
	AMP	AMC	TZP	CXM	CTX	CAZ	CRO	FEP	IPM	MEM	ETP	CIP	LEV	GEN	AMK	TGC	SXT
R1	16	8	≤8	>16	>32	2	>32	16	4	≤0.5	≤0.5	2	1	>8	>32	≤0.25	≤1
R2	≤8	≤4	≤8	>16	>32	2	>32	8	1	≤0.5	≤0.5	>2	4	>8	>32	0.5	≤1
R3	≤8	8	≤8	>16	>32	2	>32	8	2	≤0.5	≤0.5	>2	2	>8	>32	0.5	≤1
R4	>16	>16	≤8	>16	>32	>32	>32	>32	8	>8	4	1	1	>8	>32	≤0.25	>4
R5	16	8	≤8	>16	>32	4	>32	16	4	≤0.5	≤0.5	2	1	>8	>32	≤0.25	≤1
R6	≤8	≤4	≤8	>16	8	≤1	8	4	≤0.5	≤0.5	≤0.5	>2	2	4	16	≤0.25	≤1
R7	≤8	≤4	≤8	>16	32	2	32	16	≤0.5	≤0.5	≤0.5	>2	4	>8	>32	1	≤1
R8	>16	>16	32	>16	>32	>32	>32	>32	>8	>8	>4	>2	8	>8	>32	1	>4
R9	≤8	≤4	≤8	>16	32	2	32	8	1	≤0.5	≤0.5	>2	2	>8	>32	0.5	≤1
R10	≤8	8	≤8	>16	>32	2	>32	8	2	≤0.5	≤0.5	>2	2	>8	>32	0.5	≤1
R11	≤8	≤4	≤8	>16	>32	2	>32	8	1	≤0.5	≤0.5	>2	4	>8	>32	0.5	≤1
R12	≤8	≤4	≤8	>16	>32	2	>32	8	≤0.5	≤0.5	≤0.5	>2	4	>8	>32	0.5	≤1
R13	16	16	≤8	>16	>32	4	>32	16	2	≤0.5	≤0.5	>2	4	>8	>32	≤0.25	≤1
R14	16	8	≤8	>16	>32	4	>32	16	4	≤0.5	≤0.5	2	2	>8	>32	≤0.25	≤1

Table A3.5: A table to show the ration of UV absorbance at 390 nm to UV absorbance at 486 nm in each strain. Data were collected using FLUOstar Omega for 25 minutes of incubation.

Incubation Time (min)	Strain													
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
0	0.521	0.510	0.502	0.633	0.503	0.458	0.491	0.668	0.466	0.497	0.500	0.469	0.457	0.506
1	0.562	0.554	0.539	1.039	0.552	0.482	0.547	0.779	0.503	0.526	0.543	0.514	0.486	0.576
2	0.581	0.571	0.553	1.089	0.561	0.494	0.558	0.825	0.524	0.531	0.560	0.526	0.505	0.594
3	0.548	0.560	0.541	1.104	0.556	0.493	0.559	0.849	0.525	0.542	0.563	0.534	0.525	0.592
4	0.540	0.554	0.528	1.125	0.541	0.490	0.552	0.854	0.509	0.532	0.560	0.527	0.544	0.574
5	0.582	0.572	0.558	1.144	0.575	0.509	0.597	0.879	0.548	0.543	0.586	0.538	0.570	0.614
6	0.576	0.590	0.572	1.153	0.590	0.520	0.607	0.894	0.548	0.570	0.604	0.548	0.589	0.614
7	0.556	0.564	0.541	1.132	0.563	0.507	0.578	0.887	0.533	0.531	0.577	0.529	0.606	0.572
8	0.582	0.572	0.556	1.132	0.560	0.518	0.593	0.914	0.555	0.557	0.612	0.542	0.628	0.577
9	0.557	0.554	0.559	1.123	0.541	0.521	0.593	0.914	0.532	0.545	0.589	0.546	0.647	0.576
10	0.586	0.608	0.590	1.138	0.601	0.548	0.655	0.946	0.574	0.583	0.653	0.569	0.674	0.628
11	0.584	0.596	0.577	1.125	0.594	0.546	0.638	0.954	0.573	0.578	0.638	0.572	0.690	0.615
12	0.619	0.615	0.592	1.121	0.601	0.557	0.656	0.970	0.585	0.577	0.644	0.582	0.708	0.616
13	0.606	0.612	0.598	1.120	0.589	0.563	0.674	0.976	0.594	0.589	0.665	0.588	0.724	0.602
14	0.601	0.600	0.604	1.112	0.592	0.569	0.660	0.985	0.563	0.592	0.647	0.575	0.743	0.608
15	0.564	0.599	0.584	1.107	0.587	0.569	0.668	0.985	0.569	0.586	0.630	0.590	0.758	0.609

Table A3.5: A table to show the ration of UV absorbance at 486 nm to UV absorbance at 390 nm in each strain. Data were collected using FLUOstar Omega for 25 minutes of incubation (Continued).

Incubation Time (min)	Strain													
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
(Continued)														
16	0.597	0.617	0.589	1.104	0.572	0.572	0.680	0.996	0.591	0.596	0.678	0.584	0.778	0.599
17	0.606	0.620	0.599	1.102	0.606	0.587	0.706	1.009	0.603	0.591	0.667	0.601	0.798	0.604
18	0.619	0.637	0.629	1.111	0.615	0.593	0.724	1.031	0.620	0.625	0.701	0.622	0.819	0.637
19	0.613	0.640	0.626	1.102	0.607	0.601	0.724	1.035	0.641	0.625	0.696	0.635	0.837	0.642
20	0.625	0.654	0.628	1.104	0.603	0.607	0.740	1.050	0.650	0.632	0.708	0.638	0.858	0.638
21	0.625	0.652	0.642	1.104	0.613	0.614	0.740	1.057	0.640	0.630	0.721	0.635	0.875	0.633
22	0.635	0.660	0.659	1.108	0.632	0.624	0.758	1.069	0.669	0.643	0.735	0.656	0.896	0.681
23	0.653	0.664	0.655	1.106	0.629	0.629	0.775	1.071	0.654	0.657	0.747	0.656	0.908	0.659
24	0.641	0.659	0.642	1.094	0.597	0.626	0.765	1.082	0.671	0.642	0.741	0.651	0.919	0.631
25	0.644	0.676	0.663	1.099	0.652	0.642	0.789	1.086	0.661	0.669	0.770	0.662	0.946	0.665

Table A3.6: Output from CARD-enriched OrthoMCL's group

ID	Annotation
ORTHOMCL0	RND efflux system, inner membrane transporter CmeB
ORTHOMCL1	AcrB AcrD AcrF family protein
ORTHOMCL2	Cobalt zinc cadmium resistance protein CzcA, Cation efflux system protein CusA
ORTHOMCL3	RND efflux system, outer membrane lipoprotein CmeC
ORTHOMCL4	Transcriptional regulatory protein OmpR
ORTHOMCL5	ABC transporter, ATP binding permease protein
ORTHOMCL7	Chloramphenicol acetyltransferase
ORTHOMCL8	RND efflux system, membrane fusion protein CmeA
ORTHOMCL9	Transport ATP binding protein CydCD
ORTHOMCL10	Fosmidomycin resistance protein
ORTHOMCL11	Lipid A export ATP binding permease protein MsbA
ORTHOMCL12	Topoisomerase IV subunit B
ORTHOMCL13	Inner membrane component of tripartite multidrug resistance system
ORTHOMCL14	Dihydropteroate synthase
ORTHOMCL15	RND efflux system, membrane fusion protein CmeA
ORTHOMCL16	Translation elongation factor Tu
ORTHOMCL17	DNA gyrase subunit B
ORTHOMCL18	RND multidrug efflux transporter, Acriflavin resistance protein
ORTHOMCL19	Macrolide export ATP binding permease protein MacB
ORTHOMCL20	RND multidrug efflux transporter, Acriflavin resistance protein
ORTHOMCL21	RND efflux system, membrane fusion protein CmeA
ORTHOMCL22	Lipid A export ATP binding permease protein MsbA
ORTHOMCL23	DNA gyrase subunit A
ORTHOMCL24	Lipid A export ATP binding permease protein MsbA
ORTHOMCL25	Translation elongation factor G
ORTHOMCL26	DNA directed RNA polymerase beta subunit
ORTHOMCL27	RND efflux system, outer membrane lipoprotein CmeC
ORTHOMCL29	Topoisomerase IV subunit A
ORTHOMCL30	MFS permease
ORTHOMCL31	Putative ABC iron siderophore transporter, fused permease and ATPase domains
ORTHOMCL32	Lipid A export ATP binding permease protein MsbA
ORTHOMCL33	Putative streptomycin phosphotransferase
ORTHOMCL34	Membrane fusion component of tripartite multidrug resistance system
ORTHOMCL35	RND efflux system, outer membrane lipoprotein CmeC
ORTHOMCL36	RND efflux system, outer membrane lipoprotein CmeC
ORTHOMCL37	Inner membrane component of tripartite multidrug resistance system
ORTHOMCL38	RND efflux system, outer membrane lipoprotein CmeC
ORTHOMCL39	Cyclolysin secretion ATP binding protein
ORTHOMCL40	Tetracycline efflux protein TetA
ORTHOMCL41	Putative transmembrane efflux protein
ORTHOMCL43	RND efflux system, membrane fusion protein CmeA
ORTHOMCL44	RND efflux system, membrane fusion protein CmeA

ORTHOMCL45	Membrane fusion protein MexC
ORTHOMCL46	Transport ATP binding protein CydCD
ORTHOMCL47	Dihydropteroate synthase
ORTHOMCL48	ABC type multidrug transport system ATPase and permease component
ORTHOMCL49	Streptomycin-6-kinase
ORTHOMCL50	Streptomycin resistance protein B
ORTHOMCL51	Tetracycline resistance protein
ORTHOMCL52	Aminoglycoside-6' N-acetyltransferase
ORTHOMCL53	Beta lactamase class C and other penicillin binding proteins
ORTHOMCL54	Beta lactamase
ORTHOMCL55	Beta lactamase class C and other penicillin binding proteins
ORTHOMCL56	Metallo-beta lactamase superfamily protein PA0057
ORTHOMCL57	Beta lactamase
ORTHOMCL58	Metallo-beta lactamase family protein
ORTHOMCL59	Oxacillinase OXA-114
ORTHOMCL60	Beta lactamase
ORTHOMCL61	Beta lactamase domain-containing protein
ORTHOMCL62	Beta lactamase
ORTHOMCL63	Metallo-beta lactamase family protein, RNA specific
ORTHOMCL64	Beta lactamase domain containing protein
ORTHOMCL65	Beta lactamase domain protein
ORTHOMCL66	Metallo-beta lactamase

Table A3.7: Acquired resistance genes and the presence/absence in each strain

	Group number	Reference gene ID	Gene annotation	NCIMB 11015	DSM 2402	NH44784-1996	L1	L5	L8	L10	L11	L14	L15	L16	L17	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Beta-lactamase	ORTHOMCL53	AHY18660.1	IMP-14 metallo beta lactamase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-
	ORTHOMCL72	CDI94943.1	Beta lactamase domain containing protein	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	ORTHOMCL71	YP_008031839.1	Metallo beta lactamase family protein, RNA specific	-	-	+	+	+	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
	ORTHOMCL70	YP_008027752.1	Beta lactamase	-	-	+	+	-	-	-	-	-	-	-	-	+	-	+	-	+	-	-	+	-	+	-	-	+	+
	ORTHOMCL69	AIR52250.1	Beta lactamase domain containing protein	-	+	-	-	+	-	+	-	-	-	+	-	+	+	-	+	+	+	-	+	+	-	+	-	+	+
	ORTHOMCL68	YP_008029689.1	Beta lactamase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	+	+	-	+	+	+	+
Efflux transport system	ORTHOMCL1	WP_020925327.1	AcrB/AcrD/AcrF family protein	+	+	+	+	-	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
	ORTHOMCL2	CKG72640.1	Cation efflux system protein CzcA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+
	ORTHOMCL27	YP_008027732.1	RND efflux system,outer membrane CmeC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
	ORTHOMCL30	YP_008028726.1	MFS permease	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	-
	ORTHOMCL31	YP_008030221.1	Putative ABC iron siderophore transporter, fused permease and ATPase domains	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	+	-	+	+	+	+	-	+	+	-	+
	ORTHOMCL32	CKH86978.1	Putative multidrug export ATP-binding/permease protein SAV1866	-	+	+	+	+	-	+	+	-	+	-	-	+	+	-	+	+	+	-	+	+	+	+	+	+	+
	ORTHOMCL34	YP_008030767.1	Membrane fusion component of tripartite multidrug resistance system	+	+	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	-	-	+	-	-	+	+	+	-
	ORTHOMCL35	YP_008032982.1	RND efflux system outer membrane lipoprotein CmeC	-	-	+	+	+	-	+	+	+	+	+	+	+	+	-	-	+	+	-	+	+	-	+	-	+	+
	ORTHOMCL36	YP_008027734.1	RND efflux system outer membrane lipoprotein CmeC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	+	-	-	-	+	-	-
	ORTHOMCL37	YP_008033107.1	Inner membrane component of tripartite multidrug resistance system	+	+	+	+	-	+	+	-	+	-	+	+	-	-	+	+	-	+	+	+	-	-	-	+	+	-
	ORTHOMCL38	YP_008029977.1	RND efflux system outer membrane lipoprotein CmeC	+	+	+	+	-	+	+	+	+	-	+	+	+	+	-	+	+	-	-	+	-	-	-	-	-	-
	ORTHOMCL39	YP_008032049.1	Cyclolysin secretion ATP binding protein	-	-	+	+	-	+	-	-	-	-	-	+	-	+	-	-	-	+	+	+	-	-	-	+	-	-
	ORTHOMCL40	YP_008031167.1	Tetracycline efflux protein TetA	+	+	+	+	-	+	+	-	+	-	-	+	-	+	-	+	-	-	-	+	-	-	-	-	-	-
	ORTHOMCL41	CKI18085.1	Spectinomycin tetracycline efflux pump	+	-	+	+	-	+	-	-	+	-	+	+	-	-	-	+	-	-	-	+	-	-	+	+	-	-

	ORTHOMCL43	YP_008033425.1	RND efflux system membrane fusion protein CmeA	+	-	+	+	-	+	+	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	+
	ORTHOMCL44	YP_008028035.1	RND efflux system membrane fusion protein CmeA	-	-	+	+	-	+	-	-	-	-	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+
	ORTHOMCL45	ADP16465.1	Periplasmic linker protein 2	-	-	-	-	-	-	+	-	+	-	-	-	-	+	-	-	-	+	-	+	+	-	-	+	-	-
	ORTHOMCL46	YP_008031197.1	Transport ATP binding protein CydCD	+	-	-	-	-	-	+	+	-	-	+	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-
	ORTHOMCL48	CKG97466.1	Iron import ATP-binding/permease protein IrtB	-	+	-	-	-	+	-	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-
	ORTHOMCL51	AKE04698.1	MFS Transporter	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-
	ORTHOMCL54	YP_008032981.1	RND multidrug efflux transporter Acriflavin resistance protein	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	-	+	+
	ORTHOMCL55	YP_008027731.1	RND efflux system, inner membrane transporter CmeB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	+	+	-	+	+	+	+
	ORTHOMCL58	YP_008031293.1	RND efflux system, outer membrane lipoprotein CmeC	+	+	+	+	+	+	+	+	+	-	+	+	+	+	-	+	-	+	+	+	+	+	+	+	-	+
	ORTHOMCL59	WP_013393781.1	Acriflavine resistance protein B	-	-	-	-	-	-	+	-	+	-	-	-	-	+	+	-	-	+	+	+	+	+	-	+	-	-
	ORTHOMCL60	YP_008028715.1	AcrB/AcrD/AcrF family protein	+	+	+	+	+	-	+	+	-	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
	ORTHOMCL61	YP_008032003.1	RND multidrug efflux transporter; Acriflavin resistance protein	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	ORTHOMCL62	WP_020925545.1	Cobalt-zinc-cadmium resistance protein CzcA	+	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+	+	+	-	+	-	-	+	+	-	-
	ORTHOMCL63	WP_019396923.1	Cation transporter	-	-	-	-	-	-	+	+	-	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	+
Resistance-associated enzyme	ORTHOMCL16	YP_008029342.1	Translation elongation factor Tu	+	+	+	+	-	+	-	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+
	ORTHOMCL25	YP_008029360.1	Translation elongation factor G	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+
	ORTHOMCL26	YP_008029349.1	DNA directed RNA polymerase beta subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
	ORTHOMCL29	YP_008030789.1	Topoisomerase IV subunit A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+
	ORTHOMCL33	YP_008031164.1	Putative streptomycin phosphotransferase	+	-	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+	-	+	-	-	+	+	+	-
	ORTHOMCL47	AHY18660.1	Dihydropteroate synthase	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-
	ORTHOMCL49	BAN89135.1	Streptomycin-6-kinase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-
	ORTHOMCL50	AIA53578.1	3'-kinase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-
	ORTHOMCL52	AHY18660.1	Aminoglycoside-6',N-acetyltransferase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-

+, Presence; -, Absence

Table A3.8: The appearance of AxyXY-OprZ-homologous CmeABC efflux pump across 26 genomes of *A. xylosoxidans*. NH44784-1996, *A. xylosoxidans* NH44784-1996, NCIMB 11015, *A. xylosoxidans* NCIMB 11015; DSM 2402, *A. xylosoxidans* DSM 2402; L-strain, Liverpool's clinical isolates; R-strain, Thai clinical isolates.

Gene annotation	CmeA(AxyX)	CmeB(AxyY)	CmeC(OprZ)
Ref ID	YP_008027730.1	YP_008027731.1	YP_008027732.1
NCIMB 11015			
DSM 2402			
NH44784-1996			
L1			
L5			
L8			
L10			
L11			
L14			
L15			
L16			
L17			
R1			
R2			
R3			
R4			
R5			
R6			
R7			
R8			
R9			
R10			
R11			
R12			
R13			
R14			

 = Presence; = Absence

Table A3.9: The appearance of AxyAMB-homologous CmeABC efflux pump across 26 genomes of *A. xylosoxidans*. NH44784-1996, *A. xylosoxidans* NH44784-1996, NCIMB 11015, *A. xylosoxidans* NCIMB 11015; DSM 2402, *A. xylosoxidans* DSM 2402; L-strain, Liverpool's clinical isolates; R-strain, Thai clinical isolates.

Gene annotation	CmeA(AxyA)	CmeB(AxyB)	CmeC(AxyM)
Ref ID	YP_008031291.1	YP_008031292.1	YP_008031293.1
NCIMB 11015			
DSM 2402			
NH44784-1996			
L1			
L5			
L8			
L10			
L11			
L14			
L15			
L16			
L17			
R1			
R2			
R3			
R4			
R5			
R6			
R7			
R8			
R9			
R10			
R11			
R12			
R13			
R14			

 = Presence; = Absence

Table A3.10: Antibiotic susceptibility of type strains of genus *Achromobacter* to commonly used antibiotics using disc diffusion. The table presents susceptibility interpretation using cut-off level based on BSAC guideline. The susceptibility interpretation was illustrated as 'Resistant (R, Red)', 'Intermediate (I, Yellow)' and 'Susceptible (S, Green)'. NCIMB 11015, *A. xylosoxidans* NCIMB 11015; DSM 2402, *A. xylosoxidans* DSM 2402

	Cephalosporin					Aminoglycoside	
Isolates	CXM	CTX	CAZ	CRO	FEP	CN	AK
NCIMB 11015	R	R	S	R	I	R	I
A. denitrificans	R	I	S	S	I	R	I
A. ruhlandii	R	R	S	I	S	S	S
A. piechaudii	R	R	S	R	S	S	S
A. spanius	R	R	S	I	S	S	S
A. insolitus	R	R	S	R	R	I	S
DSM 2402	R	R	S	R	R	R	R

Appendix 4 Virulence determinants of *A. xylosoxidans*

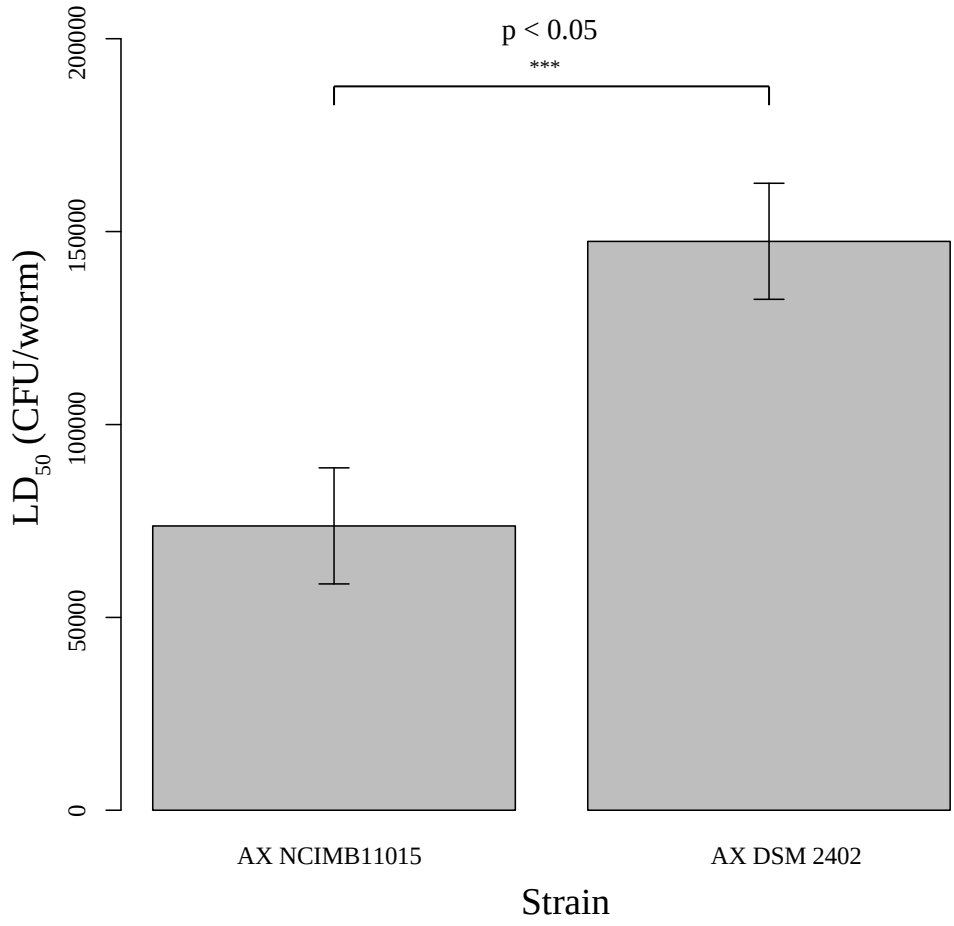


Figure A4.1: Lethal dose-50 (LD_{50}) of *A. xylosoxidans* strains in *G. mellonella*. The larvae were infected with two *A. xylosoxidans* strains: NCIMB 11015 and DSM 2402, and incubated for 96 hr at 37°C. LD_{50}s were determined using R programme. Bars represent the means and standard deviation of three replicates. Student's t-test was used as statistical test. ***; p-value < 0.005.

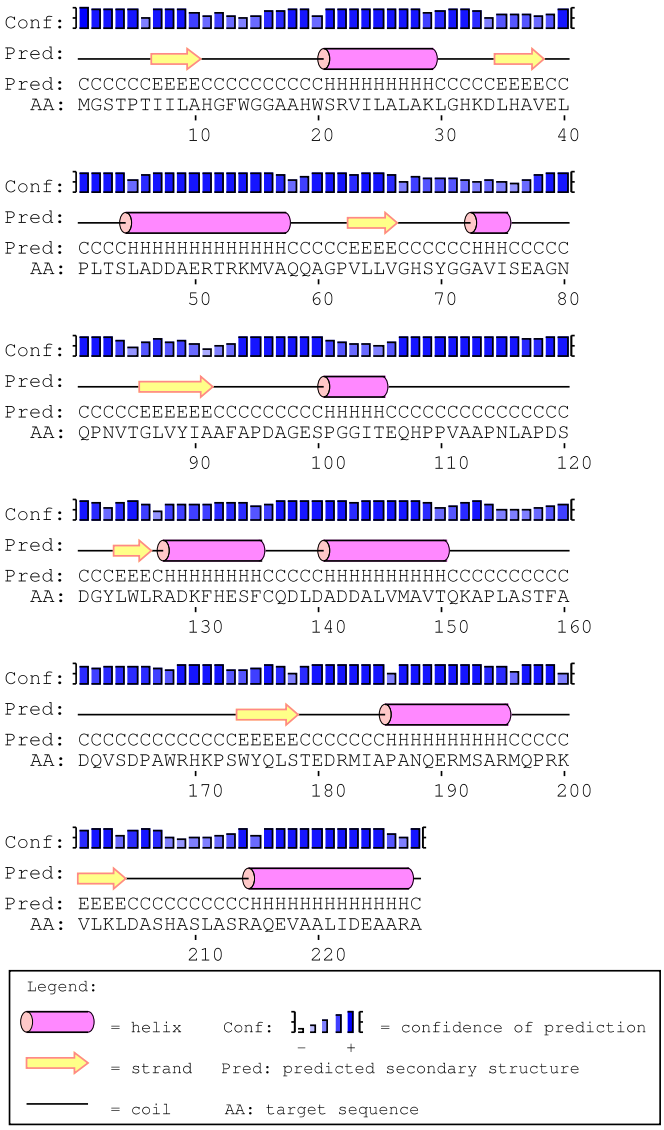


Figure A4.2: The prediction of secondary structure of virulence-associated α/β -hydrolase domain-containing protein. The analysis was performed using Psipred (http://bioinf.cs.ucl.ac.uk/psipred/).

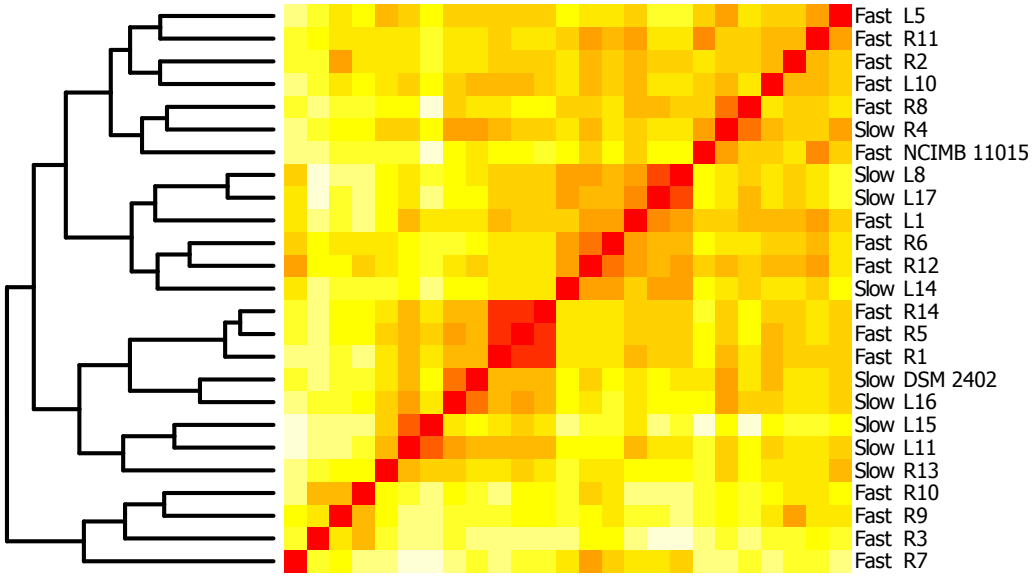


Figure A4.3: Heatmap presenting hierarchical cluster of *A. xylosoxidans* isolates based on virulence gene candidates. A dendrogram represents the clustering by using UPGMA method. Colours on the heatmap indicate the correlation between isolates based on the occurrence of VFDB-matched virulence gene. Lethality and isolates' name were labelled on the right-hand side of the heatmap. NCIMB 11015, *A. xylosoxidans* NCIMB 11015; DSM 2402, *A. xylosoxidans* DSM 2402. L-strain, Liverpool's clinical isolates; R-strain, Thai clinical isolates.

Table A4.1: A list of VFDB-enriched OrthoMCL's groups that are shared amongst *A. xylosoxidans* isolates in this study

ID	Annotation
ORTHOMCL0	General secretion pathway protein E
ORTHOMCL1	Heat shock protein 60 family chaperone GroEL
ORTHOMCL2	Outer membrane receptor for ferric pyochelin FptA
ORTHOMCL3	Phospholipase C4 precursor
ORTHOMCL4	Iron siderophore Alcaligin receptor/TonB-dependent siderophore receptor
ORTHOMCL5	Arginine decarboxylase
ORTHOMCL6	RND efflux system outer membrane lipoprotein CmeC
ORTHOMCL7	Isocitrate lyase
ORTHOMCL8	Co activator of prophage gene expression IbrA
ORTHOMCL9	ABC transporter ATP binding permease protein
ORTHOMCL10	GDP mannose-4,6-dehydratase
ORTHOMCL11	Mannose-1-phosphate guanylyltransferase GDP
ORTHOMCL12	UDP glucose 4 epimerase
ORTHOMCL13	Succinate semialdehyde dehydrogenase NADP
ORTHOMCL14	ClpB protein
ORTHOMCL15	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase
ORTHOMCL16	Superoxide dismutase Fe
ORTHOMCL17	ATP dependent Clp protease ATP binding subunit ClpA
ORTHOMCL18	ATP dependent Clp protease proteolytic subunit
ORTHOMCL19	Lysyl tRNA synthetase class II
ORTHOMCL20	Methyl accepting chemotaxis protein I serine chemoreceptor protein
ORTHOMCL21	putative protein ImpC
ORTHOMCL22	Methyl accepting chemotaxis protein I serine chemoreceptor protein
ORTHOMCL23	Sulfite oxidase
ORTHOMCL24	Two-component response regulator
ORTHOMCL25	Respiratory nitrate reductase beta chain
ORTHOMCL26	2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase I alpha
ORTHOMCL27	Vi polysaccharide export protein vexE
ORTHOMCL28	Vi polysaccharide biosynthesis protein tviE
ORTHOMCL29	putative capsular polysaccharide biosynthesis protein
ORTHOMCL30	UDP glucose dehydrogenase
ORTHOMCL31	Glutamine synthetase type I
ORTHOMCL32	Flagellar M ring protein FliF
ORTHOMCL33	Flagellar motor switch protein FliG
ORTHOMCL34	Flagellum specific ATP synthase FliI

ORTHOMCL35	Flagellar motor switch protein FliM
ORTHOMCL36	Flagellar biosynthesis protein FliP
ORTHOMCL37	Flagellar hook associated protein FlgK
ORTHOMCL38	Flagellar protein FlgJ peptidoglycan hydrolase
ORTHOMCL39	Flagellar P ring protein FlgI
ORTHOMCL40	Flagellar basal body rod protein FlgG
ORTHOMCL41	Flagellar biosynthesis protein FlhB
ORTHOMCL42	Chemotaxis protein methyltransferase CheR
ORTHOMCL43	Positive regulator of CheA protein activity CheW
ORTHOMCL44	Signal transduction histidine kinase CheA
ORTHOMCL45	Flagellar motor rotation protein MotB
ORTHOMCL46	RNA polymerase sigma factor for flagellar operon
ORTHOMCL47	Aldehyde dehydrogenase
ORTHOMCL48	Transport ATP binding protein CydCD
ORTHOMCL49	Tetrathionate reductase subunit B
ORTHOMCL50	Biofilm PGA synthesis N glycosyltransferase PgaC
ORTHOMCL51	Lipid A export ATP binding permease protein MsbA
ORTHOMCL52	Putative ABC transporter
ORTHOMCL53	Phenylacetaldehyde dehydrogenase
ORTHOMCL54	C4 dicarboxylate transport transcriptional regulatory protein
ORTHOMCL55	Putative ATP binding component of ABC transporter
ORTHOMCL56	Hemin transport protein HmuS
ORTHOMCL57	ATP dependent Clp protease proteolytic subunit
ORTHOMCL58	2-aminomuconate-semialdehyde dehydrogenase
ORTHOMCL59	UDP-glucose 4-epimerase
ORTHOMCL60	UDP-glucose dehydrogenase
ORTHOMCL61	TonB-dependent siderophore receptor
ORTHOMCL62	Polymyxin resistance protein ArnC glycosyl transferase
ORTHOMCL63	DNA binding heavy metal response regulator
ORTHOMCL64	Aldehyde dehydrogenase
ORTHOMCL65	Polymyxin resistance protein ArnC glycosyl transferase
ORTHOMCL66	Pyruvate kinase
ORTHOMCL67	Gamma-glutamyltranspeptidase
ORTHOMCL68	Mg 2 transport ATPase P type
ORTHOMCL69	UTP-glucose-1-phosphate uridylyltransferase
ORTHOMCL70	TonB-dependent siderophore receptor
ORTHOMCL71	RNA polymerase sigma factor RpoE
ORTHOMCL72	Lipid A export ATP binding permease protein MsbA
ORTHOMCL73	ADP-L-glycero-D-manno-heptose 6-epimerase
ORTHOMCL74	2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase I alpha

ORTHOMCL75	Lipid A export ATP binding permease protein MsbA

Appendix 5 Complete genome of *Achromobacter xylosoxidans* strain NCIMB 11015, a clinical-naïve strain

The genome of clinically-naïve strain of *A. xylosoxidans* included in this study was revealed with the combination between short read sequencing and single molecule sequencing technology. Genome assembly presented 6.5 Mb genome long, with 67.6% of G+C content. Rapid gene calling and genes annotation with Prokka annotator resulted in 5,864 CDS from the genome with 1,072 hypothetical genes. There were 71 tRNA genes and 3 rRNA operons predicted on the genome.

With functional annotation of genome based on eggNOG database (Appendix Table A5.1), approximately one-fourth of genes were assigned as 'function unknown'. Besides genes with unknown function, genes associated with amino acid metabolism/transport and gene associated with transcription ranked the most and the second most abundant genes in the genome with 12 percent and 9.8 percent, respectively. In comparison to average of COGs assigned in other strains, *A. xylosoxidans* NCIMB 11015 had less number of genes annotated with 'poor characterized'.

Whole genome comparison of the genome of NCIMB 11015 to the other genome was performed with BLAST-based comparison and graphical view was generated using BRIG software (Appendix Figure A5.1). Using visualised comparative genomic approach, there were 8 regions that are unique to strain NCIMB 11015, as illustrated by black boxes in Appendix Figure A5.1 and genes in each region were described in Appendix Table 5.2. The first region (Box I in Appendix Figure A5.1) was located at 1,495,202 bp to 1,513,912 bp on the genome. Most of the genes in this region were hypothetical genes. The presence of a gene encoding 'integrase' at the end of the segment suggested the acquisition of this DNA fragment from environment (Appendix Table A5.2). The second region spanned from 1,710,796 to 1,730,209 bp (Box II in Appendix Figure A5.1). Genes in this region were also mostly hypothetical genes.

UvrD/REP helicase and precursor of type III restriction (Res subunit) were present in this DNA region. The third region (Box III in Appendix Figure A5.1), which located at position approximately 2.2 Mbp, was composed of genes associating with Lipopolysaccharide (LPS) biosysnthesis, such as, LPS-heptosyltransferase and 3-deoxy-D-manno-octulosonic acid transferase. In addition, genes involved in bacterial capsule biosynthesis, including UDP-glucose 4-epimerase and UDP-N-acetylglucosamine 2-epimerase, were present in this region. Genes with incomplete sequence or pseudo-genes of transposase were also found in this region. The fourth region (Box IV in Appendix Figure A5.1), was comprised of components of type IV secretion system, including inner membrane protein and minor pilin. Moreover, this region included major facilitator, transcription regulators, kinase and integrase family protein. The fifth region (Box V in Appendix Figure A5.1), located on genome spanning from coordination 3,388,522 to 3,444,827, included type III secretion system's subunits and integrase/terminase genes. The sixth and the largest region covered 77,774 bp (3,819,279 – 3,897,053) on the genome (Box VI in Appendix Figure A5.1). This region consisted of genes involved in flagella, two-component systems, pilus and secretion system's components. In region VII (Box VII in Appendix Figure A5.1), most of the genes were hypothetical genes. This region could be phage-related due to the presence of phage protein. The last region (Box VIII in Appendix Figure A5.1) was composed of only single gene, which is hypothetical protein.

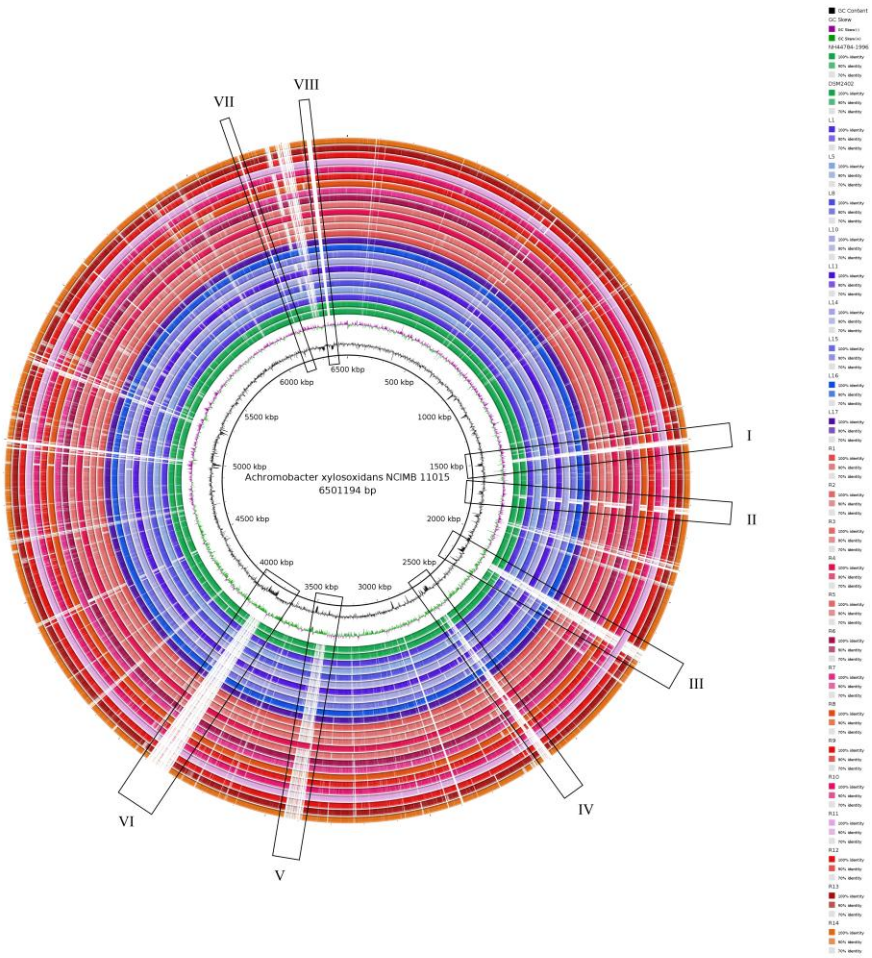


Figure A5.1: A circular genome map comparing *A. xylosoxidans* NCIMB 11015 to the other isolates of *A. xylosoxidans* used in this study. Eight black boxes indicate regions where are unique to strain NCIMB 11015.

Table A5.1: The functional annotation of complete genome of *A. xylosoxidans* strain NCIMB 11015 compared to other 25 *A. xylosoxidans* genomes

Categories	COGs function	NCIMB 11015	Mean of 25 strains
Information storage and processing	(A) RNA processing and modification	0.002	0.02%
	(B) Chromatin structure and dynamics	0.1	0.09%
	(J) Translation	3.1	2.84%
	(K) Transcription	9.8	7.91%
	(L) Replication, recombination, repair	2.4	2.27%
Cellular processes	(D) Cell cycle control, mitosis, meiosis	0.5	0.49%
	(M) Cell wall/membrane biogenesis	4.5	4.19%
	(N) Cell motility	0.8	0.75%
	(O) Posttranslational modification, protein turnover, chaperones	3.0	2.68%
	(T) Signal transduction mechanisms	3.5	3.12%
	(U) Intracellular trafficking, secretion	1.7	1.33%
	(V) Defense mechanisms	1.3	0.97%
	(W) Extracellular structures	0.0	0.01%
	(Z) Cytoskeleton	0.0	0.00%
Metabolism	(C) Energy production, conversion	6.2	4.87%
	(E) Amino acid transport, metabolism	12.0	9.70%
	(F) Nucleotide transport, metabolism	1.2	1.03%
	(G) Carbohydrate transport, metabolism	3.1	12.98%
	(H) Coenzyme transport, metabolism	1.9	1.85%
	(I) Lipid transport, metabolism	4.1	3.51%
	(P) Inorganic ion transport, metabolism	6.4	5.55%
	(Q) Secondary metabolites	1.8	3.84%
Poorly characterised	(R) General function prediction	6.0	5.02%
	(S) Function unknown	15.1	24.96

Table A5.2: A table to show unique genomic regions of *A. xylosoxidans* NCIMB 11015 compared to other strains.

Region I	Start	Stop
Conjugal Transfer Protein	1,495,202	1,495,816
Hypothetical Protein	1,495,903	1,496,466
Hypothetical Protein	1,496,268	1,496,657
Hypothetical Protein	1,496,667	1,497,119
Outer Membrane Porin Protein Precursor	1,497,142	1,497,948
Hypothetical Protein	1,498,214	1,499,188
Hypothetical Protein	1,499,317	1,502,292
Hypothetical Protein	1,502,945	1,503,229
Putative Transcriptional Regulator	1,503,213	1,503,533
Parb Domain Protein Nuclease	1,503,652	1,504,212
Hypothetical Protein	1,504,475	1,505,377
Hypothetical Protein	1,506,782	1,510,327
Integrase Family Protein	1,512,743	1,513,912
Region II		
Putative Type III Restriction Protein, Res Subunit	1,710,796	1,713,393
Hypothetical Protein	1,713,390	1,715,387
Hypothetical Protein	1,715,384	1,716,721
Hypothetical Protein	1,716,841	1,717,791
Hypothetical Protein	1,717,887	1,718,834
Uvrd/REP Helicase	1,718,827	1,722,030
Hypothetical Protein	1,722,033	1,724,726
Hypothetical Protein	1,724,799	1,726,454
Putative Transcriptional Regulator	1,727,234	1,728,259
Hypothetical Protein	1,728,388	1,729,116
Hypothetical Protein	1,729,586	1,730,209
Region III		
Transposase (Pseudogene?)	2,164,425	2,164,742
Transposase (Pseudogene?)	2,164,736	2,165,050
Transposase (Pseudogene?)	2,165,051	2,165,230
Carbon Monoxide Degydrogenase G Protein	2,165,347	2,165,916
Hypothetical Protein	2,166,489	2,168,213
Hypothetical Protein	2,169,073	2,169,846
Hypothetical Protein	2,169,882	2,170,880
Degt/Dnrj/Eryc1/Strs	2,170,877	2,172,094
Hypothetical Protein	2,172,091	2,173,647
Hypothetical Protein	2,173,671	2,174,936
Hypothetical Protein	2,174,957	2,176,666
Dolichol-Phosphate Mannosyltransferase	2,176,859	2,177,665
UDP-N-Acetylglucosamine 2-Epimerase	2,177,955	2,178,746
Hypothetical Protein	2,179,365	2,180,216
Multidrug MFS Transporter	2,180,741	2,182,579
Acetyltransferase	2,182,599	2,183,198

Glycosyl Transferase	2,183,252	2,184,283
NAD-Dependent Epimerase/Dehydratase Family Protein 7	2,184,341	2,185,222
Glycosyl Transferase Family 1	2,185,219	2,186,472
UDP-N-Acetylglucosamine 2-Epimerase	2,186,454	2,187,587
UDP-2-Acetamido-2,6-Dideoxy-Beta-L-Talose 4 Dehydrogenase	2,187,589	2,188,695
UDP-Glucose 4-Epimerase	2,188,698	2,189,735
Imidazole Glyceraol Phosphate Synthase	2,189,774	2,190,559
Imidazole Glyceraol Phosphate Synthase	2,190,561	2,191,175
LPS Biosynthesis Protein Wbpg	2,191,172	2,192,311
Glycosyl Transferase	2,192,346	2,193,275
Hypothetical Protein	2,193,658	2,194,986
Polysaccharide Biosynthesis Protein (O-Antigen-Associated)	2,194,997	2,196,223
Aminotransferase Degt	2,196,361	2,197,473
Serine Acetyltransferase	2,197,473	2,198,060
Oxidoreductase	2,198,057	2,198,950
Hypothetical Protein	2,199,640	2,200,710
LPS-Heptosyltransferase	2,201,112	2,202,140
3-Deoxy-D-Manno-Octulosonic Acid Transferase	2,202,142	2,203,443
Region IV		
Hypothetical Protein	2,588,317	2,588,919
Addiction Module Antidote Protein	2,589,058	2,589,360
Peptidase C14, Caspase Catalytic Subunit P20	2,589,725	2,590,822
Hypothetical Protein	2,590,847	2,593,336
Hypothetical Protein	2,593,341	2,594,180
Hypothetical Protein	2,594,184	2,594,969
Hypothetical Protein	2,595,140	2,596,435
Integral Inner Membrane Protein Of Type IV Secretion Complex, Virb6	2,597,641	2,598,693
Hypothetical Protein	2,598,695	2,598,955
Hypothetical Protein	2,598,952	2,599,398
Minor Pilin Of Type IV Secretion Complex, Virb5	2,599,409	2,600,134
Hypothetical Protein	2,600,392	2,600,649
Hypothetical Protein	2,600,675	2,601,670
Hypothetical Protein	2,601,667	2,602,206
Hypothetical Protein	2,602,203	2,603,213
Phosphoprotein Phosphatase	2,603,904	2,604,593
Alkylhydroperoxidase	2,605,471	2,606,010
Major Facilitator Superfamily	2,606,079	2,607,272
Short-Chain Dehydrogenase/Reductase	2,607,330	2,608,088
Arac Family Transcriptional Regulator	2,608,217	2,609,167
Hypothetical Protein	2,611,601	2,611,726
Integrase Family Protein	2,611,849	2,612,475
Alkylhydroperoxidase	2,612,645	2,613,091
Transcriptional Regulator, Gntr Family Domain/ Aspartate Aminotransferase	2,613,248	2,614,666
Putative Kinase	2,614,909	2,616,003
Hypothetical Protein	2,616,000	2,616,428

Region V		
Hypothetical Protein	3,388,522	3,388,611
Hypothetical Protein	3,388,637	3,389,155
Hypothetical Protein	3,389,209	3,390,186
Hypothetical Protein	3,390,868	3,391,515
Hypothetical Protein	3,392,054	3,393,040
Hypothetical Protein	3,393,522	3,393,653
Terminase	3,393,798	3,394,067
Hypothetical Protein	3,394,398	3,394,667
Integrase Family Protein	3,394,621	3,395,319
Hypothetical Protein	3,395,876	3,397,282
Hypothetical Protein	3,397,358	3,398,707
Hypothetical Protein	3,398,704	3,398,826
Hypothetical Protein	3,398,911	3,400,080
Srcm	3,400,107	3,400,565
Putative Inner Membrane Transporter Protein	3,400,562	3,403,057
ABC Transporter, Substrate-Binding Protein	3,403,258	3,404,421
Tir Chaperone Protein	3,404,703	3,405,146
Hypothetical Protein	3,405,319	3,406,368
Hypothetical Protein	3,406,449	3,406,910
Hypothetical Protein	3,406,932	3,408,209
Hypothetical Protein	3,408,310	3,409,290
Hypothetical Protein	3,409,512	3,409,763
Hypothetical Protein	3,409,804	3,410,061
Putative Type III Secretion Protein	3,410,091	3,411,413
Type III Secretion Inner Membrane Channel Protein	3,411,410	3,413,503
Hypothetical Protein	3,413,496	3,413,864
Hypothetical Protein	3,413,861	3,414,259
Hypothetical Protein	3,414,298	3,414,666
Putative Outer Protein B	3,414,663	3,415,808
Inositol Phosphate Phosphatase	3,415,983	3,418,691
Hypothetical Protein	3,418,736	3,419,206
Secreted Protein 22	3,419,312	3,419,911
Type III Secretion Chaperone For Yopd	3,420,262	3,420,762
Hypothetical Protein	3,420,773	3,421,675
Putative Outer Protein B	3,421,713	3,422,942
Hypothetical Protein	3,419,913	3,420,218
Putative Regulator Protein	3,422,970	3,423,455
Hypothetical Protein	3,423,574	3,423,882
Putative Type III Secretion Protein	3,423,891	3,424,325
Type III Secretion Bridge Between Inner And Outer Membrane Lipoprotein	3,424,325	3,425,179
Bsck	3,425,176	3,425,826
Type III Secretion Cytoplasmic Protein	3,425,796	3,426,434
Flagellum-Specific ATP Synthase Fli 1	3,426,431	3,427,798
Hypothetical Protein	3,428,317	3,428,892

Type III Secretion Spans Bacterial Envelope Protein	3,427,812	3,428,330
Hypothetical Protein	3,428,914	3,430,089
Hypothetical Protein	3,430,086	3,430,778
Hypothetical Protein	3,430,807	3,431,073
Hypothetical Protein	3,431,093	3,431,893
Hypothetical Protein	3,431,890	3,432,951
Hypothetical Protein	3,432,938	3,433,354
Type III Secretion Outer Membrane Pore Forming Protein	3,433,351	3,435,345
RNA Polymerase ECF-Type Sigma Factor	3,435,380	3,435,985
Response Regulator, Narl Family	3,436,323	3,436,994
ABC Transporter Substrate Protein	3,437,355	3,438,575
Serine Phosphatase Rsbu, Regulator Of Sigma Subunit	3,438,572	3,440,578
Putative Anti-Sigma Factor	3,440,766	3,441,206
Putative Anti-Sigma Factor Antagonist	3,441,357	3,441,716
Hypothetical Protein	3,441,918	3,443,339
Hypothetical Protein	3,443,469	3,444,827
Region VI		
Hypothetical Protein	3,819,279	3,820,289
Hypothetical Protein	3,822,085	3,822,789
Hypothetical Protein	3,823,862	3,825,661
Flagellar Transcriptional Regulator	3,825,827	3,826,375
Hypothetical Protein	3,827,419	3,828,396
Putative Two-Component Response Regulator	3,829,378	3,830,238
Flagellar Transcriptional Activator	3,826,403	3,826,738
Hypothetical Protein	3,830,379	3,830,747
Multi- Hybrid Histidine Kinase	3,831,126	3,834,236
Luxr-Family DNA-Binding Response Regulator	3,834,284	3,834,937
Glutamate-Aspartate Periplasmic Binding Proterin Precursor	3,835,007	3,836,032
Periplasmic Thiol:Disulfide Interchange Protein	3,836,191	3,836,802
CS1 Type Fimbrial Major Subunit	3,836,874	3,837,401
Putative Periplasmic Chaperone	3,837,507	3,838,220
Putative Outer Membrane Usher	3,838,318	3,841,134
Putative Minor Pilin And Initiator	3,841,257	3,842,336
Glycoside Hydrolase Family	3,842,594	3,845,857
Type II Secretion System Protein G	3,846,224	3,846,664
Invasion Protin Iagb	3,846,639	3,847,151
Type II Secretion System Protein	3,847,148	3,848,332
Type II Secretion System Protein E	3,848,405	3,850,009
Hypothetical Protein	3,849,990	3,850,763
Hypothetical Protein	3,850,760	3,851,290
Hypothetical Protein	3,851,287	3,851,841
Hypothetical Protein	3,851,838	3,852,308
Bacterial Type II And III Secretion System Amily Protein	3,852,312	3,854,327
General Secretion Pathway Family Protein G	3,854,324	3,854,824
Type II Secretion System Protein G	3,854,799	3,855,194
Hypothetical Protein	3,855,181	3,855,681

Two-Component Transcriptional Regulator, Winged Helix Family	3,855,720	3,856,403
Putative Transmembrane Sensor Domain Protein	3,856,410	3,857,714
Histidine Kinase A Phosphoacceptor Domain-Containing Protein	3,857,711	3,858,784
Type III Secretion System Protein	3,858,741	3,859,907
Ysat	3,859,904	3,860,686
Putative Type III Secretion 318Ystem Protein	3,860,686	3,860,940
Epap	3,860,953	3,861,645
Type III Secretion Apparatus Protein	3,861,642	3,862,574
Hypothetical Protein	3,862,589	3,864,451
Hypothetical Protein	3,864,441	3,864,905
Type III Secretion System Atpase	3,864,892	3,866,151
Hypothetical Protein	3,866,193	3,866,603
Inva	3,866,611	3,868,680
Hypothetical Protein	3,868,714	3,869,034
Type Secretion System Regulator	3,869,027	3,870,157
Type III Secretion Apparatus	3,870,150	3,871,985
Hypothetical Protein	3,871,976	3,872,275
Transcriptional Regulator	3,872,253	3,873,032
Hypothetical Protein	3,873,210	3,873,350
Hypothetical Protein	3,873,558	3,874,730
Hypothetical Protein	3,874,850	3,875,131
Hypothetical Protein	3,875,159	3,875,455
Hypothetical Protein	3,875,452	3,876,210
Hypothetical Protein	3,876,940	3,877,467
Hypothetical Protein	3,877,559	3,878,659
Chaperone Protein	3,878,893	3,879,390
Hypothetical Protein	3,879,383	3,881,341
Hypothetical Protein	3,881,409	3,882,311
Hypothetical Protein	3,882,395	3,885,583
Hypothetical Protein	3,885,790	3,886,077
Hypothetical Protein	3,886,608	3,888,113
Hypothetical Protein	3,888,209	3,889,687
Hypothetical Protein	3,889,851	3,890,489
Hypothetical Protein	3,890,523	3,891,545
RNA 2'-Phosphotranferase	3,891,908	3,892,735
Hypothetical Protein	3,892,926	3,897,053
Region VII		
Hypothetical Protein	6,282,837	6,283,502
Hypothetical Protein	6,283,506	6,283,694
Hypothetical Protein	6,283,783	6,284,076
Hypothetical Protein	6,284,086	6,284,304
Hypothetical Protein	6,284,301	6,284,828
Glycosyl Hydrolase	6,284,825	6,285,319
Hypothetical Protein	6,285,316	6,285,648
Hypothetical Protein	6,285,781	6,286,167
Hypothetical Protein	6,286,170	6,287,210

Hypothetical Protein	6,287,211	6,287,861
Hypothetical Protein	6,287,858	6,289,048
Hypothetical Protein	6,289,041	6,289,385
Hypothetical Protein	6,289,382	6,290,089
Hypothetical Protein	6,290,074	6,290,898
Hypothetical Protein	6,290,891	6,291,184
Hypothetical Protein	6,291,196	6,291,738
Peptidoglycan-Binding Protein	6,291,738	6,293,354
Hypothetical Protein	6,293,445	6,293,900
Hypothetical Protein	6,293,910	6,294,359
Hypothetical Protein	6,294,418	6,295,926
Hypothetical Protein	6,295,936	6,296,463
Hypothetical Protein	6,296,460	6,296,828
Hypothetical Protein	6,296,825	6,297,301
Hypothetical Protein	6,297,298	6,297,726
Hypothetical Protein	6,297,728	6,298,078
Hypothetical Protein	6,298,138	6,299,220
Hypothetical Protein	6,299,236	6,299,718
Hypothetical Protein	6,299,715	6,300,980
Phage Head Protein	6,300,961	6,301,713
Head Protein	6,301,742	6,303,277
Phage Terminase, Large Subunit	6,303,274	6,304,395
Terminase	6,304,499	6,304,918
Hypothetical Protein	6,304,969	6,305,265
Hypothetical Protein	6,305,262	6,305,543
Hypothetical Protein	6,305,606	6,305,815
Hypothetical Protein	6,305,952	6,306,320
DNA Helicase	6,306,321	6,307,709
Hypothetical Protein	6,307,706	6,308,551
Hypothetical Protein	6,308,551	6,308,916
Hypothetical Protein	6,309,006	6,309,416
Phage-Related Transcriptional Regulator	6,309,416	6,309,730
Hypothetical Protein	6,309,769	6,309,933
Hypothetical Protein	6,309,936	6,310,199
Putative Phage Repressor Protein	6,310,276	6,311,373
Hypothetical Protein	6,311,443	6,312,246
Hypothetical Protein	6,312,847	6,313,503
Hypothetical Protein	6,314,034	6,314,243
Hypothetical Protein	6,314,318	6,314,497
Pentapeptide Repeat Protein	6,314,781	6,315,374
Hypothetical Protein	6,315,388	6,315,924
Hypothetical Protein	6,316,737	6,316,937
Hypothetical Protein	6,317,590	6,317,679
Hypothetical Protein	6,317,681	6,317,953
Hypothetical Protein	6,317,956	6,318,462
Hypothetical Protein	6,318,459	6,318,644

Hypothetical Protein	6,318,683	6,319,336
Single-Stranded DNA-Binding Protein	6,319,336	6,319,779
Hypothetical Protein	6,319,788	6,320,291
Hypothetical Protein	6,320,377	6,320,523
Hypothetical Protein	6,320,525	6,321,073
Hypothetical Protein	6,321,076	6,321,309
Hypothetical Protein	6,321,313	6,321,495
Hypothetical Protein	6,321,492	6,322,457
Hypothetical Protein	6,322,450	6,323,115
Hypothetical Protein	6,323,108	6,323,323
Hypothetical Protein	6,323,323	6,323,613
Hypothetical Protein	6,323,890	6,324,063
Hypothetical Protein	6,324,072	6,324,293
Integrase Family Protein	6,324,269	6,325,444
Region VIII		
Hypothetical Gene	6,378,002	6,392,161

Appendix 6 Additional recombination analysis

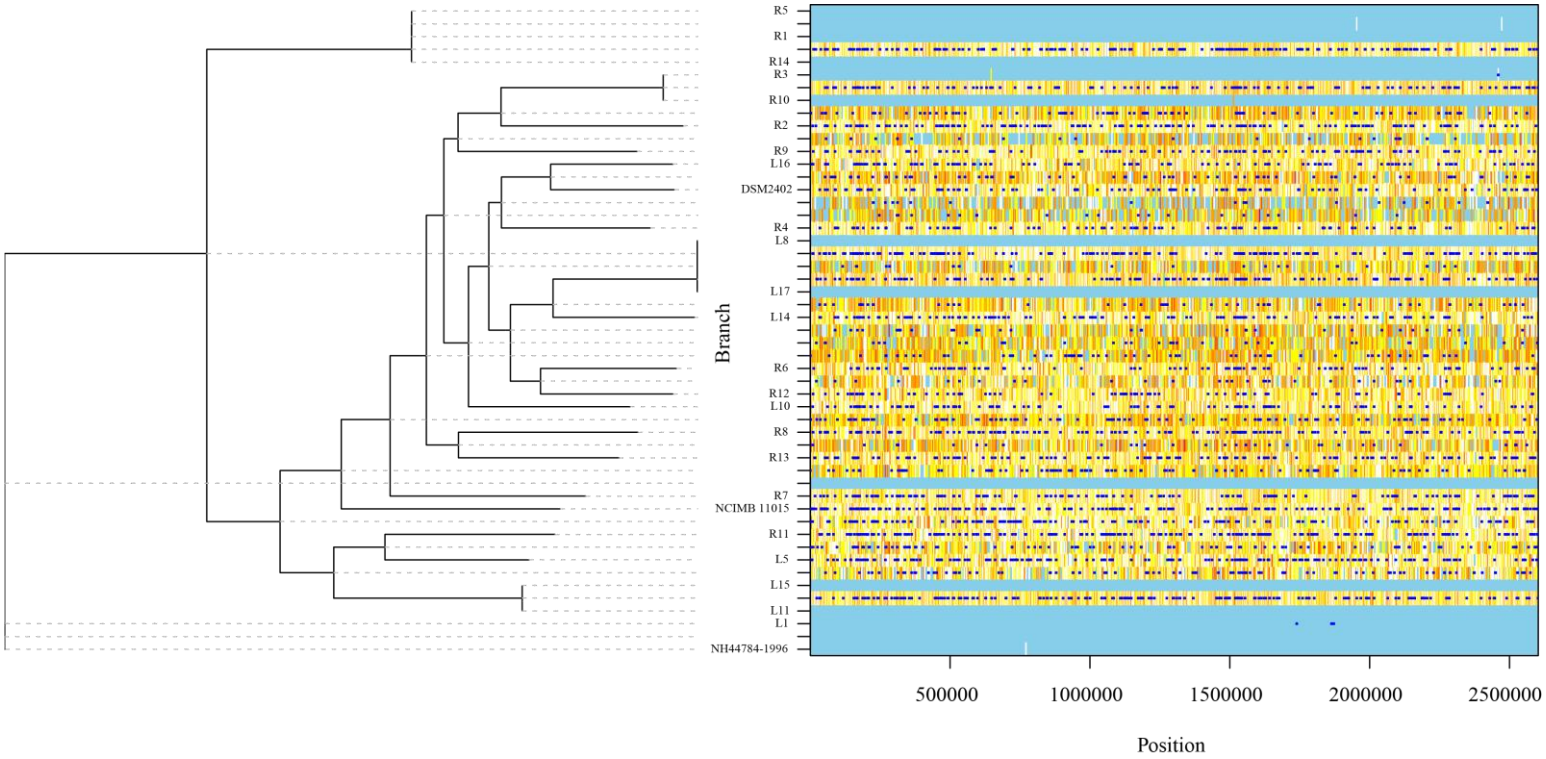


Figure A6.1: Recombination analysis on the core genomes of 26 *A. xylosoxidans* isolates using ClonalframML. Dark blue dots illustrate recombination detected by the software

Figure A6.2: A plot of recombination distribution across 21 *B. hensalae* genomes. A PSA tree is shown on the right side and the names of the strains are shown on the left side of the plot. Colour of each bar in a plot indicates recombined sequences obtained from a corresponding cluster in a dendogram.